DOCUMENT RESUME

ED 253 395 SE 045 358

AUTHOR	Bethel, Lowell J., Ed.
TITLE	Research and Curriculum Development in Science Education. 4. Curriculum Evaluation, Classroom Methodology, and Theoretical Models. The University of Texas Centennial Science Education Center Monograph.
INSTITUTION	Texas Univ., Austin. Science Education Center.
PUB DATE	84
NOTE	181p.
AVAILABLE FROM	Science Education Center, The University of Texas at Austin, Austin, TX 78712 ($12.95).
PUB TYPE	Reports - Research/Technical (143) -- Collected Works - General (020)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Classroom Techniques; *Curriculum Development; Curriculum Evaluation; Elementary Secondary Education; Environmental Education; Higher Education; *Inservice Teacher Education; Instructional Material Evaluation; Methods Courses; *Models; Preservice Teacher Education; Process Education; Science Curriculum; *Science Education; *Science Instruction; Second Language Learning; *Teaching Methods
IDENTIFIERS	Misconceptions; *Science Education Research

ABSTRACT

This monograph consists of 17 research studies in science education. Among the areas investigated in these studies are: modeling and monitoring of college level science education; classroom processes and their relationships with short-term outcomes in elementary school science; expectations of the scientific enterprise; a science laboratory safety computer program simulation for preservice teachers; a minicourse on emotional stress; effects of science inquiry activities on the development of classification and communication skills of children learning a second language; a delphi study of the objectives and planning concerns for secondary science methods courses; influence of class ability level on biology teaching practices; evaluation of an environmental science education academic year program for intermediate elementary teachers (4-6); effectiveness of student feedback and self-analysis in changing selected science teaching practices; effect of rating-delay, instructional content, and length of attendance of elementary teacher ratings of science inservice workshops; effectiveness of teacher-prepared handouts in learning science concepts; detecting and classifying the origins of science misconceptions; effect of a persuasive communication on attitudes toward energy use of preservice elementary teachers; and a comparison of methods for measuring the degree of implementation of an innovative science program. (JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED253395

RESEARCH AND CURRICULUM DEVELOPMENT IN SCIENCE EDUCATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SE 045 358

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
James P. Barufaldi
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Copyright ©1984 by the Science Education Center
THE UNIVERSITY OF TEXAS AT AUSTIN

*Copies of this publication may
be procured for $12.95 from the
SCIENCE EDUCATION CENTER,
THE UNIVERSITY OF TEXAS AT AUSTIN,
AUSTIN, TEXAS 78712.*

FOREWORD

The Science Education Center released its first research monograph in October, 1967. The monograph was a collection of research reports based on work completed by both faculty and graduate students in the Center at The University of Texas at Austin. This is the fourth monograph in the series.

The fourth monograph is a very special one. It is special because of it being released during the university's centennial celebration. Many projects and activities have been undertaken during 1983 and 1984 to celebrate this milestone in the university's long and honored history. The faculty of the Science Education Center feel that this is an appropriate way to celebrate the centennial.

The monograph is being released at a time when science education is on the cutting edge of a national debate initiated by the publication of three major reports: *Science & Engineering Education for the 1980's & Beyond, Educating Americans for the 21st Century*, and *A Nation at Risk.*

The research reported here focuses on a broad spectrum of problems and related issues in science education. Areas of interest include curriculum evaluation, classroom methodology, program evaluation, and theoretical models. All of these areas are critical to science education and the research reports provide some important insight and suggest future directions for research.

Many people have contributed time and much effort to the development of this university centennial research monograph. Many thanks go to Dr. Rolland B. Bartholomew, Dr. John P. Huntsberger, Dr. Thomas R. Koballa, Jr., Dr. Frank Crawley, and Dr. Earl J. Montague who critiqued the research reports included in the monograph. I am also grateful for the work and support of many graduate students on this research monograph including Delena Tull, Glen W. Harral, and Kathleen A. O'Sullivan. Appreciation is extended to Lynn Fox, Pat Taylor, Laura Avant, Valerie Milton, Mrs. Bonner Wilhelm, and Mrs. Agnes E. Bunn who were all responsible for typing and preparing the manuscript for publication; and special thanks is extended to the Director of the Science Education Center, Dr. James P. Barufaldi, whose unstinting support and help made possible the publication of this monograph.

LOWELL J. BETHEL
Austin, Texas
March, 1984

The search for the truth is in one way hard and in another easy — for it is evident that no one of us can ever master it fully, nor miss it wholly. Each one of us adds a little to our knowledge of nature and from all the facts assembled arises a certain grandeur.

Aristotle, 350 B.C.

Sit down before fact as a little child, be prepared to give up every preconceived notion, follow humbly wherever and to whatever abyss nature leads, or you shall learn nothing. . . .

T. H. Huxley, 1942 A.D.

TABLE OF CONTENTS

I. Modeling and monitoring college level science education . . . Karen Johnston 1

II. Classroom processes and their relationships with short term outcomes in elementary school science . . . Charles W. Anderson 16

III. Expectations of the scientific enterprise . . . Susan Helms 28

IV. A science laboratory safety computer program simulation for preservice teachers . . . Floyd D. Ploeger 37

V. Teaching the most basic skill — our survival depends on it . . . Marianne Betkouski 44

VI. Effects of science inquiry activities on the development of classification and communication skills of children learning a second language . . . Imelda Z. Rodriguez 52

VII. Objectives and planning concerns for secondary science methods courses: A delphi study . . . Marcia Virginia dos Santos Silva 64

VIII. The influence of class ability level on biology teaching practices . . . Julie P. Sanford 75

IX. Evaluation of an environmental science education academic year program for intermediate elementary teachers (4-6) . . . Lowell J. Bethel and Shirley M. Hord 88

X. The effectiveness of pupil feedback and self-analysis in changing selected science teaching practices . . . Melanie Lewis and Rolland Bartholomew 100

XI. The effect of rating-delay, instructional content, and length of attendance on elementary teacher ratings of science inservice workshops . . . Thomas C. Meadows 110

XII. The effect of a persuasive communication on the attitude toward energy use of preservice elementary teachers . . . Thomas R. Koballa, Jr. 124

XIII. A comparison of methods for measuring the degree of implementation of an innovative science program . . . James D. Ellis 132

XIV. An investigation of the relationship between student characteristics and cognitive gains obtained by undergraduate students in an audio-visual tutorial self-paced science program . . . Robert M. Holland 143

XV. Detecting and classifying the origins of science misconceptions . . . Richard James Trembath 152

XVI. The effects of participation in selected activities from the biological science curriculum study-elementary school sciences program on the enhancement of listening skills . . . Jennifer R. W. Swift 162

XVII. On the effectiveness of teacher-prepared handouts in learning science concepts . . . James A. Petrich 171

DEVELOPMENT AND ANALYSIS OF A MODEL TO DESCRIBE THE STATE OF SCIENCE EDUCATION AT THE COLLEGE LEVEL IN TEXAS

KAREN L. JOHNSTON

DEPARTMENT OF PHYSICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA

INTRODUCTION

Traditionally, the functions of decision making and planning in higher education have been left to institutional administrators and state legislators. Often these decisions and plans for action have been made with limited and/or incomplete information about existing educational activities. In other words, decisions that affect various aspects of higher education such as financing, curriculum changes, and meeting manpower needs, have been made without the availability of a descriptive or quantitative model of the existing educational endeavor (Correa, 1975). This deficiency in the information system on which important decisions are based could result in courses of action that do not satisfy local, state, and national goals and needs (Henle, 1967).

PROBLEM

Since 1972 the National Science Board (NSB) has attempted to monitor science and technology in this country. The work of the Board is published periodically in a document titled *Science Indicators*. The document contains statistical data that reflect the health of science and technology in the United States. One area not included in *Science Indicators* is the area of science education. Science education is an important support system for science and technology. Clearly, the need exists to monitor periodically science education as the NSB has attempted to do with other aspects of science and technology. Data such as this can be used by decision-makers in order to develop and prepare programs that contribute to the nation's science and technology efforts. But there appears to be no systematic system for determining the health and status of science education through time.

PURPOSE

States are increasingly being called upon to do something about the critical shortages of science teachers and poorly trained students in science. Funds originally allocated to NSF and other educational programs have been returned to the states in the form of block grants. But if state educational planners are to find new programs to meet these needs and requirements critical data are required. The study described in this paper was an initial attempt to develop a model for assessing the health of science education at a statewide level. Such a model could then be used by state planners for program development and decision-making. In addition, the model developed could be utilized for national assessment.

RESEARCH QUESTIONS

The study focused on a small sector of public higher education: science education at the college level at public institutions in the state of Texas. The following research questions were addressed:

1. What are the major functions of science education at the college level in Texas?
2. What are the measures (data through time or *time series*) that describe science education at the college level?
3. What numerical indexes can be used to monitor science education at the college level for the State of Texas?

OPERATIONAL DEFINITIONS

As used in this study specific terms are defined as follows:

SCIENCE EDUCATION:	the activities and endeavors related to education in the physical and biological science disciplines
TIME SERIES DATA:	a set of observations taken over a period of time
INDICATOR:	time series data that are useful in describing the status of science education
INDEX NUMBER:	a measurement that is conceptually useful in describing the status of science education

RESEARCH DESIGN

The utilization of models is widespread in the natural sciences. The purpose of a model is to represent reality. Models for complex systems are useful because the models can: (a) provide a reference frame for consideration of a problem, (b) permit abstraction (from the real world to the model world) that can be used to support decision making, and (c) enhance communication through the use of symbols (Bross, 1953). Since science education is a complex system, the task of describing the status of this system requires the development of a model of science education.

A diagram of the descriptive model of science education at the college level developed in this investigation is shown in Figure 1. The intent of the descriptive model is to: (a) identify the basic "variables" (called *functions* in this paper), (b) collect information about these basic variables (called *indicators* or *time series*), and (c) descriptively and quantitatively analyze the information collected in order to draw conclusions about the status of science education at the college level (see Rains, 1979). Knowledge about the science education endeavor is achieved through descriptions of the basic variables that characterize it.

PROCEDURE

To answer the research questions and achieve the purpose of the study, the procedure was divided into three distinct phases that: (a) identified the functions of science education via questionnaire input from science educators, (b) identified time series that describe these functions of science education via questionnaire input, and (c) analyzed the information to reveal those measures that would be most useful in determining trends and in monitoring science education in Texas. The procedural approach of this study is illustrated in Figure 2.

The objective of the first research question was to identify the functions of science education at the college level and to select a specified number (five) for detailed study. To answer these questions an intensive literature search was undertaken to determine the functions of science education at the college level. Nine functions were identified and are shown in Table A. The College Science Education Questionnaire (CSEQ) was developed and administered to science educators in Texas. Part I of the CSEQ listed the nine functions of science education identified from the literature and requested that the respondents rank order the functions from "most important" to "least important." A single category of "Other" permitted the respondents to identify any other major function of science education that was not included in the nine previously identified functions. From this rank ordering, the five most important functions were selected for detailed study (see Table A).

To address the second research question, Part II of the CSEQ was designed to identify those statistical measures or time series data that might serve as indicators of the status of the functions of science education identified in Part I of the CSEQ. The respondents were requested to answer in free response style the question, "What are the measures (time series data) that are recommended for statistical study of the nine previously identified functions of science education at the college level?"

The purpose of research question #3 was to identify and describe special purpose index numbers that could be used to monitor the status of science education through time. This identification was accomplished by descriptively analyzing the time series data. The suggested numerical indexes for monitoring science education through time were determined in large part by the type and availability of time series data.

INSTRUMENTATION

The College Science Education Questionnaire (CSEQ) was designed to identify five primary functions of science education at the college level and to suggest numerical data available through time that would describe the status of these functions. CSEQ was pilot tested for errors and ambiguities prior to its distribution to the respondents. Respondent profile data were collected and results indicated that the average length of time that respondents had in science education and/or college education was 20 years. Eighty percent of the respondents held doctorate degrees. Fifty percent described their activities as primarily teaching while 25 percent spent most of their time in administrative tasks. Of the respondents

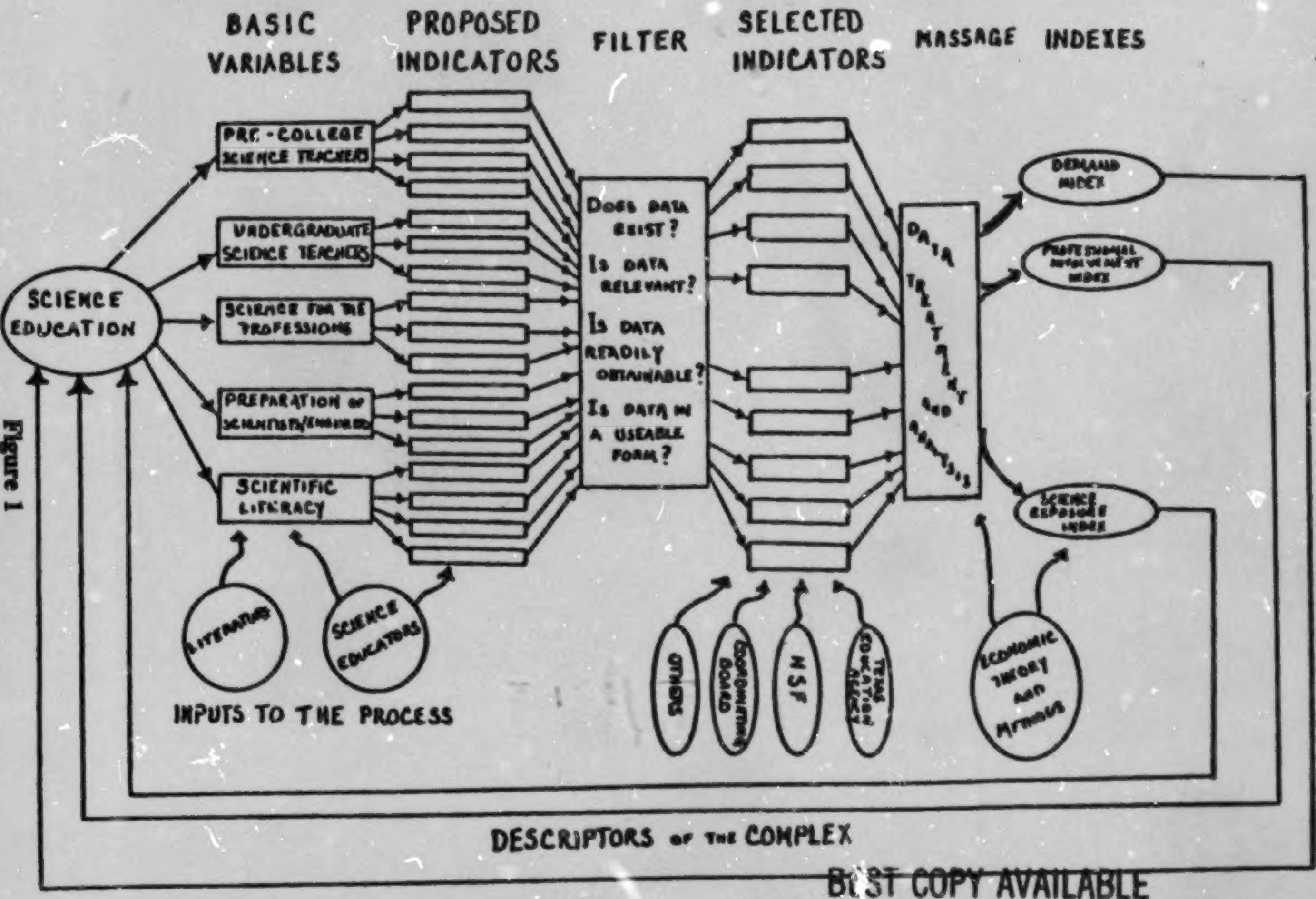

BASIC VARIABLES
PROPOSED INDICATORS
FILTER
SELECTED INDICATORS
MASSAGE
INDEXES
SCIENCE EDUCATION
PRE-COLLEGE SCIENCE TEACHERS
UNDERGRADUATE SCIENCE TEACHERS
SCIENCE FOR THE PROFESSIONS
PREPARATION OF SCIENTISTS/ENGINEERS
SCIENTIFIC LITERACY
Does data exist?
Is data relevant?
Is data readily obtainable?
Is data in a useable form?
DATA TREATMENT AND ANALYSIS
DEMAND INDEX
PROFESSIONAL INVOLVEMENT INDEX
SCIENCE EXPOSURE INDEX
LITERATURE
SCIENCE EDUCATORS
INPUTS TO THE PROCESS
OTHERS
COORDINATING BOARD
NSF
TEXAS EDUCATION AGENCY
ECONOMIC THEORY AND METHODS
DESCRIPTORS OF THE COMPLEX

Figure 1
DESCRIPTIVE MODEL OF SCIENCE EDUCATION

BEST COPY AVAILABLE

who originally agreed to participate in the study, 70 percent actually completed the CSEQ.

From Part I of the questionnaire the nine functions of science education at the college level that were rank ordered, the top five functions were selected

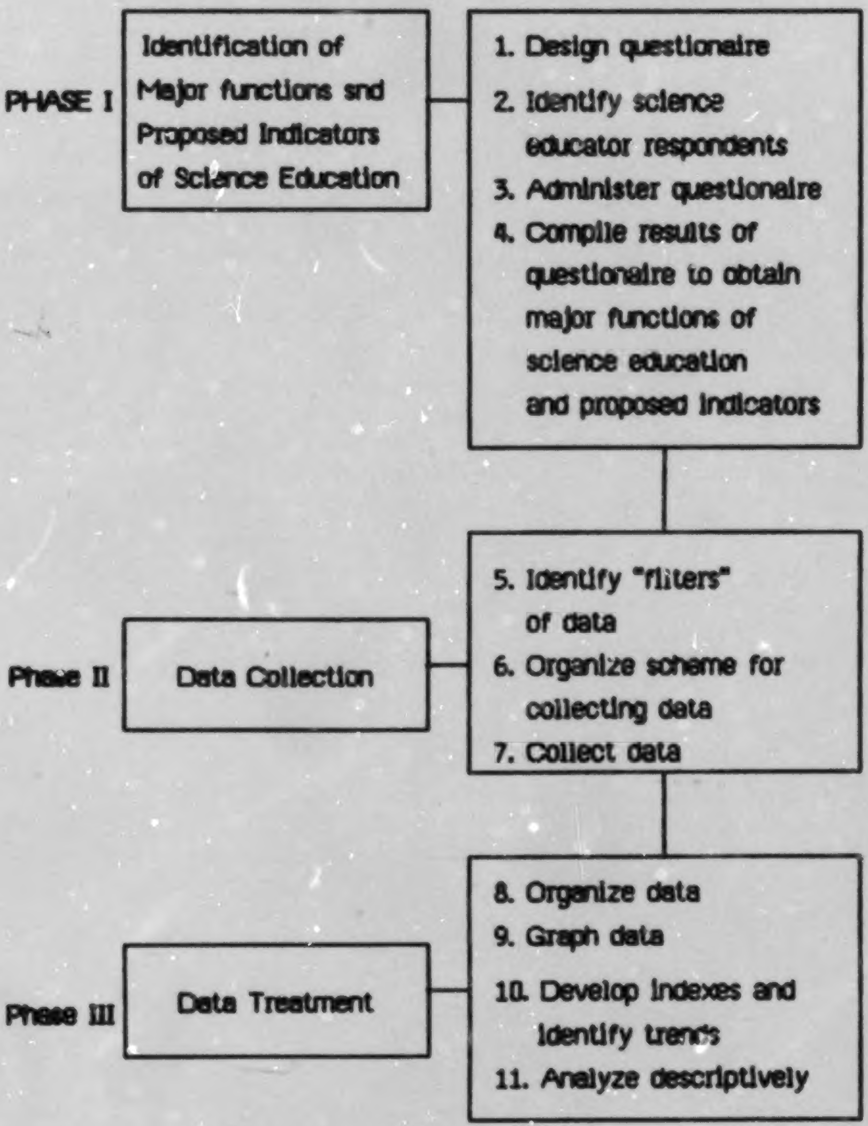

PHASE I
Identification of Major functions snd Proposed Indicators of Science Education
1. Design questionaire
2. Identify science educator respondents
3. Administer questionaire
4. Compile results of questionaire to obtain major functions of science education and proposed indicators
Phase II
Data Collection
5. Identify "filters" of data
6. Organize scheme for collecting data
7. Collect data
Phase III
Data Treatment
8. Organize data
9. Graph data
10. Develop indexes and identify trends
11. Analyze descriptively

Figure 2

PROCEDURAL APPROACH TO THE RESEARCH PROBLEM

for detailed study. These are identified in Table A. The responses in Part II of the CSEQ were lists of proposed numerical data that might be collected in order to analyze the status of that particular function of science education. These responses were filtered through a series of questions aimed at ascertaining the appropriateness of the suggested time series as a measure of the function and the availability of the data. The proposed indicators (time series data) that survived the filter questions are discussed in the following section.

SCIENCE EDUCATION INDICATORS AND HIGHLIGHTS

Each of the following subsections consists of a presentation and discussion of the time series data related to the function of science education. Not all series data that were collected were presented in the interest of space. Data that illustrate the method utilized in arriving at suggested numerical indexes were presented and highlights for all series data are featured.

PREPARATION OF THE PRE-COLLEGE SCIENCE TEACHER

Time series data related to teacher certification rate for the state were obtained from the Texas Education Agency. The year 1972 was selected as the base since this was a peak year for teacher certification in the state (Kidd, 1979). The data are summarized in the form of certification relatives which are each year's percentage of the base year with the base year having a value of 100 percent. Figure 3 displays certification relatives and permits visual identification of possible trends.

The data on certification relatives show: (a) a declining trend for total numbers certified to teach science, (b) a declining trend in certification for physics and chemistry since 1972, and (c) a declining trend in biology certification even though it still exceeds certification in the other science areas by a ratio of three to one.

HIGHLIGHTS

- Based on the College Science Education Questionnaire, the development of an index number that permits monitoring the QUALITY of pre-college science teachers would be extremely beneficial to the science education enterprise. The definition of the QUALITY INDEX has not been established and adequate data for the development of this index are not presently available.
- Certification of teachers in physics, chemistry and biology has declined since 1972.
- A demand for pre-college science teachers with certification in life, physical and composite sciences is evident and likely to continue (Kidd, 1978).

UNDERGRADUATE SCIENCE TEACHERS

Information on the number of advanced degrees in the science disciplines awarded by Texas colleges and universities is readily available. The series data

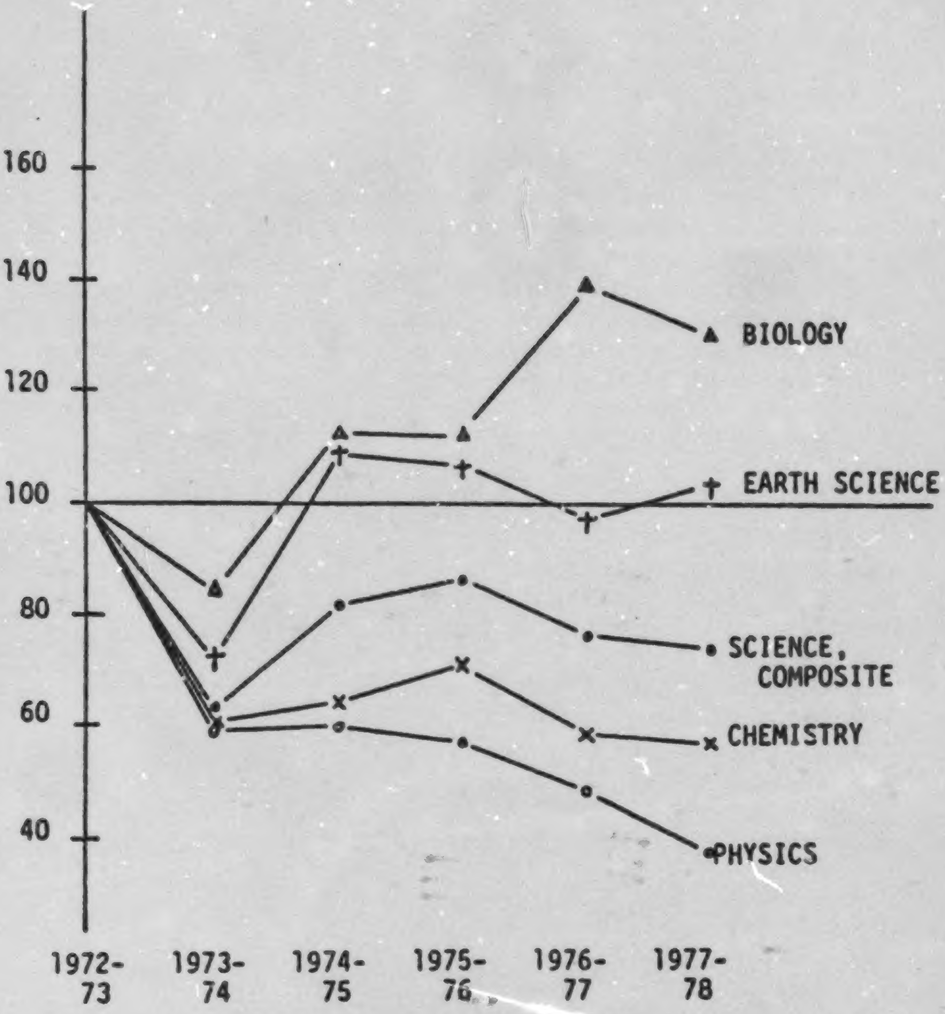


Figure 3

CERTIFICATION RELATIVES FOR SELECTED SECONDARY SCIENCE AREAS

provide information on the available pool of prospective science teachers prepared in Texas only and does not provide demand information that could be gained from monitoring numbers of new hires who are recent graduates. Figures 4 and 5 show the relatives of doctorates and master's degrees awarded in selected science areas. During the first part of the decade there was a rise in the total production of doctorates in science while the latter part of the decade indicates a leveling off of people who form a portion of the potential pool for undergraduate science teachers. Figure 5 shows that the production of master's degrees in all science disciplines combined has increased over a ten year period with the physics discipline showing a decline.

HIGHLIGHTS

- In the latter part of the decade the production of doctorate degrees in the science disciplines appears to be undergoing small fluctuations around the 1972 level.
- Doctorate production in physics has steadily increased through the decade while the production of master's degrees in physics has declined.
- Production of master's degrees in all science areas, except physics, has steadily increased through the decade.

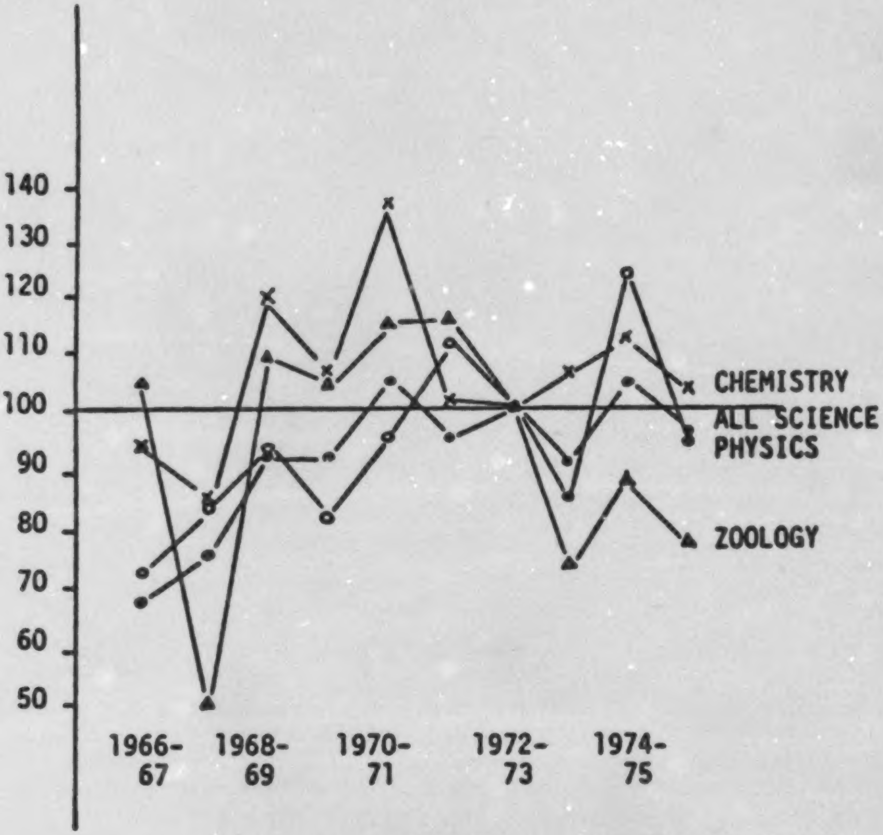


Figure 4

RELATIVES FOR DOCTORATES AWARDED IN SELECTED SCIENCE AREAS (1972, BASE YEAR)

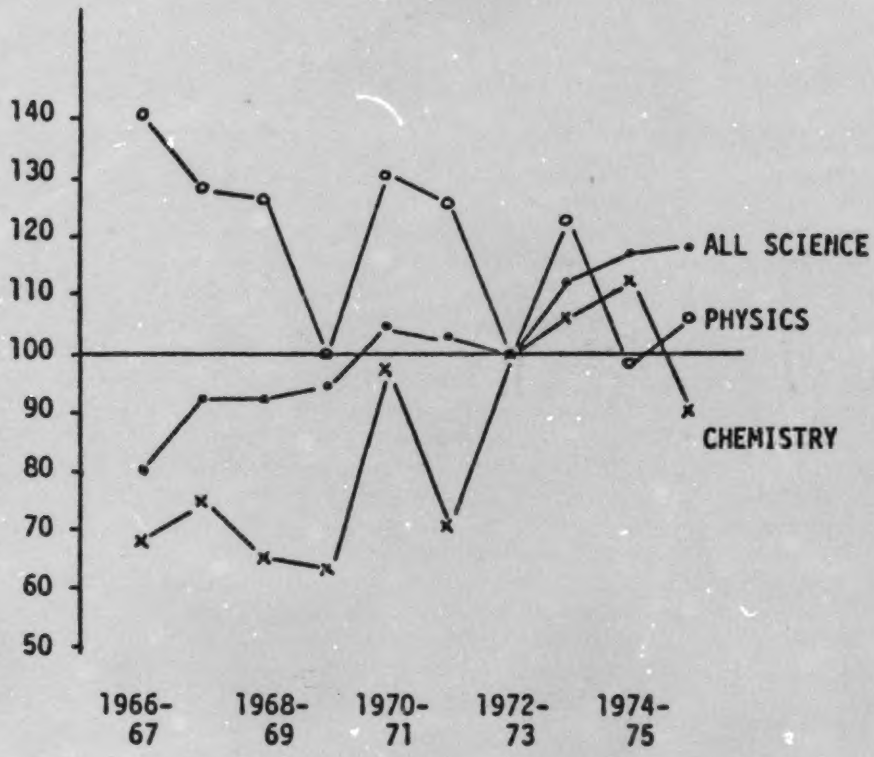


Figure 5

RELATIVES FOR MASTERS AWARDED IN SELECTED SCIENCE AREAS (1972 BASE YEAR)

- A PROFESSIONAL INVOLVEMENT INDEX is suggested for use at university departmental levels. This index could be developed from agreed upon measures of the faculty member's professional involvement. Use of such an index would, by necessity, require that departments establish identifiable standards for promotions and merit raises. Insufficient data exists to produce this index at this time.
- A DEMAND INDEX for undergraduate science teachers would be useful for counseling and graduate admittance purposes. Insufficient data on employment trends after graduation, however, preclude the development of this index at present.

SCIENCE FOR STUDENTS ENTERING THE PROFESSIONS

Since science credit requirements differ for each type of professional school and the ways in which students prepare themselves for the specific professional schools are highly varied, information on this function of science education was scarce and in a form insufficient for descriptive analysis.

PREPARATION OF SCIENTISTS AND ENGINEERS

Typically science departments at public colleges in the state do not conduct controlled follow-up studies on the employment and/or future education of their science graduates. Any data that are available from placement centers on students obtaining employment as a professional scientist are sparse.

In contrast, there is a higher degree of certainty that the engineering graduates of Texas colleges enter engineering careers. Placement centers at colleges with large engineering programs seem to do a good job of monitoring changes in numbers of engineering graduates, salary offers to graduates, and areas of industry that employ the new graduates. Figures collected from The University of Texas at Austin reveal a rapid increase in the number of new engineers entering the market. The increase can generally be interpreted as an increase in demand for engineers by industry. According to the Engineering Career Assistance Center of the University of Texas at Austin, there was a 51% increase in the number of companies interviewing engineering graduates during the 1977-78 year over the previous year. The demand for petroleum engineers continues to rise, followed by mechanical engineers and chemical engineers. Aerospace and civil engineers were the only groups experiencing decreases in salary offers. No index numbers are suggested for the statistical information related to the preparation of scientists and engineers due to both the nature and scope of the data.

HIGHLIGHTS

- No conclusions can be drawn about the balance between supply and demand for scientists prepared in Texas public institutions of higher learning at present.
- Funding received by Texas public institutions of higher education in the Undergraduate Research Participation and the Graduate Traineeship Programs of the National Science Foundation has remained healthy through the decade for which data are available.
- Petroleum, mechanical, and chemical engineering graduates (in that order) remain in demand and continue to experience substantial increases in initial salary offers.

SCIENTIFIC LITERACY OF THE GENERAL PUBLIC

No statistical data are available on the level of the general public for the state of Texas. To provide at least some "measure" of this factor, time series information was available on enrollment in science courses in Texas public schools for a 14 year period. This is illustrated in Table B. In order to analyze these data, Table C on the *Science Exposure Index* was constructed in the following manner:

Column (a) represents the fraction of 7th grade students enrolled in life science;
Column (b) represents the fraction of 8th grade students enrolled in Earth science;
Column (c) represents the fraction of 9th grade students enrolled in physical science;
Column (d) represents the fraction of 10th grade students enrolled in biology;
Column (e) represents the fraction of 11th grade students enrolled in chemistry;
Column (f) represents the fraction of 12th grade students enrolled in physics;
Column (g) represents the fraction of 12th grade students enrolled in biology II;
Column (h) represents the fraction of 12th grade students enrolled in chemistry II;
Column (i) represents the fraction of 12th grade students enrolled in physics II.

The decimal fraction of students enrolled provides information about the numbers of students in each grade level *exposed* to science in Texas public high schools. It is suggested that any combinations of these numbers can produce what will be termed a *SCIENCE EXPOSURE INDEX.*

THE *SCIENCE EXPOSURE INDEX* for all *science* areas for the 1964-65 academic year is *2.7.* This *SCIENCE EXPOSURE INDEX* is obtained by adding each of the decimal fractions that are measurements of the science exposure for each science area. That is,

SCIENCE EXPOSURE INDEX = .5645 + .6225 + .2508 + .7864 + .3689
(all areas, 1964-65) + .1494 + .0008 = 2.7436 ≈ 2.7

Or, another example, the *SCIENCE EXPOSURE INDEX* for *advanced* science courses in 1971-72 (advanced science courses defined as biology II, chemistry II and physics II) is obtained by adding each of the decimal fractions that represent the exposure of students to each of the advanced science courses. That is,

SCIENCE EXPOSURE INDEX = .1003 + .0541 + .0230 = .1774 ≈ .18
(advanced science, 1971-72)

The usefulness of this index number can be described in the following manner. Increased emphasis on science by schools or states or nations is generally evidenced by increased enrollments in public school science courses. If a school district sets increased exposure to science as a goal for its educational program, then the district can establish certain criteria that define that particular goal. This criteria might look something like the following:

80% of the students in 9th grade should be enrolled in physical science;
80% of the students in the 10th grade should be enrolled in biology;
50% of the students in the 11th grade should be enrolled in chemistry;
40% of the 12th grade students should be enrolled in physics.

With this criteria, then, the ideal *SCIENCE EXPOSURE INDEX* is 2.5. An index below 2.5 means the school has not yet met its goal while an index above 2.5 means that the school has met its goal. Since the goals will be established by the school, the "ideal" value of the *SCIENCE EXPOSURE INDEX* will vary depending on the school goal. Hence no argument need be made for the "normal value" of the *SCIENCE EXPOSURE INDEX.*

HIGHLIGHTS

- The *SCIENCE EXPOSURE INDEX* for primary science areas shows a downward trend that can be interpreted as a decrease in the general public's exposure to science.
- The *SCIENCE EXPOSURE INDEX* is suggested as a useful measure for monitoring science exposure in the public schools. It is useful because it permits schools to define precisely the emphasis they wish to place on science and determine easily whether or not they are meeting their established goal.
- Data from the Natural Science Reading sub-test of the ACT Assessment seem to indicate a trend toward declining scores.

CONCLUSIONS

The intent of this investigation was to develop a model of science education at the college level such that the science education endeavor could be monitored numerically through time. The model developed was a heuristic model in that it focused on characterizing the nature of science education and generated some prospects for index numbers that could be used to monitor the health of the basic functions of science education at the state level.

The significance of this investigation is that it attempted to integrate long standing concepts and practices used in the study of business cycles and economic activities for use in solving educational problems. An index number, when properly constructed for the purpose at hand, can reduce large, cumbersome figures into a single number that can be easily understood and followed for monitoring purposes. In addition, index numbers provide an extremely useful means for determining whether or not a school or system has met a particular goal, once the school or system has established such a goal. Perhaps the most utilitarian feature of the index number is that in order to construct the index, the purpose and the type of data needed for construction must be determined in advance. This feature alone requires *planning* by decision-makers and users of the index and eliminates the collection of useless or meaningless numerical data.

Sound educational decisions for the future must be based on some type of systematic monitoring technique rather than the opinions of experts about the health of science education. Index numbers were suggested for this use by Fisher (1927) more than 50 years ago. But it can be concluded, based on this investigation's findings, that there was little data available for trend analysis. So it is clear that educational decision-makers still have not developed clear cut methods for monitoring systematically the health of science education at the state level.

In order to plan for the future in science education, the needs of the enterprise must be anticipated. Through periodic monitoring, it is conceivable that some level of prediction of future needs can be made, with a lead time sufficient in length to intervene in order to attempt to meet future needs. In business and economics considerable work has been done in identifying series that serve as leading indicators of other business activities. Through observation of the leading indicators, predictions can be made about the related business activities and interventions. This same type of consistent and careful monitoring should become a prime concern for those involved in making decisions about science education in the state of Texas.

Table A

NINE FUNCTIONS OF SCIENCE EDUCATION RESPONSE FREQUENCIES SUMMARY*

Research in science education	7	
Communication of research findings	9	
Preparation of pre-college science teachers	20	**
Preparation of undergraduate science teachers	14	**
Curricula: pre-college	10	
Curricula: undergraduate	6	
Increasing scientific literacy	12	**
Preparation of scientists/engineers	11	**
Science for pre-professional programs	11	**

*Based on the number of responses that placed the given function in one of the top five rankings.
**Function selected for detailed study.

Table B

ENROLLMENTS IN TEXAS PUBLIC SCHOOL SCIENCE CLASSES

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1964-65	117,444	118,712	45,833	132,925	53,682	19,622	39	0	108
1965-66	163,357	146,773	87,243	146,275	56,222	17,[illegible]06	2736	1502	346
1966-67	102,894	99,915	161,420	159,482	55,338	16,056	6697	3428	2873
1967-68	125,613	94,283	168,824	164,011	54,950	15,515	7737	2873	1418
1968-69	132,539	98,006	181,888	174,782	55,823	17,257	5606	2881	1974
1969-70	144,926	111,684	188,154	178,510	53,661	18,809	6184	2765	878
1970-71	148,669	125,274	201,146	188,887	55,218	17,754	7535	3588	1129
1971-72	167,211	119,582	208,109	194,442	52,164	16,135	8696	3701	1341
1972-73	166,237	138,435	206,557	195,132	50,062	16,787	8302	3353	877
1973-74	170,906	142,264	196,587	189,423	49,807	14,421	8722	3293	1008
1974-75	152,181	129,439	171,934	162,993	41,107	11,968	7213	3000	903
1975-76	154,628	136,261	198,608	190,073	49,607	14,491	8869	3060	672
1976-77	154,230	134,921	194,407	194,407	50,450	14,848	10233	3083	853
1977-78	136,101	120,100	177,795	181,681	50,560	15,730	9300	2373	759

SOURCE: Files, Texas Education Agency

Column (a): Life Science
Column (b): Earth Science
Column (c): Physical Science
Column (d): Biology I
Column (e): Chemistry I
Column (f): Physics I
Column (g): Biology II
Column (h): Chemistry II
Column (i): Physics II

Table C
SCIENCE EXPOSURE INDEXES

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1964-65	.5645	.6225	.2508	.7864	.3689	.1494	.0003	—	.0008
1965-66	.7588	.7292	.4517	.8454	.3711	.1325	.0207	.0114	.0026
1966-67	.4657	.4783	.7909	.8457	.3562	.1178	.0492	.0252	.0211
1967-68	.5606	.4389	.7886	.8550	.3341	.1113	.0555	.0206	.0102
1968-69	.5724	.4482	.8260	.8709	.3262	.1170	.0380	.0195	.0134
1969-70	.6113	.4931	.8422	.8708	.2891	.1238	.0407	.0182	.0058
1970-71	.6317	.5413	.8662	.9058	.2984	.1109	.0471	.0224	.0071
1971-72	.7054	.5415	.9111	.9438	.2907	.1003	.0541	.0230	.0083
1972-73	.6894	.5945	.8675	.8828	.2628	.1026	.0507	.0205	.0054
1973-74	.6992	.5986	.8142	.8615	.2567	.0859	.0520	.0196	.0060
1974-75	.6376	.5572	.7273	.7693	.2223	.0729	.0439	.0183	.0055
1975-76	.6511	.5747	.8182	.8623	.2603	.0865	.0529	.0183	.0040
1976-77	.6132	.5432	.7420	.8104	.2425	.0832	.0573	.0173	.0048
1977-78	.5886	.5040	.7099	.7859	.2493	.0882	.0522	.0133	.0043

SOURCE: Files, Texas Education Agency, computed from data and Table C

Column (a): Life Science
Column (b): Earth Science
Column (c): Physical Science
Column (d): Biology I
Column (e): Chemistry I
Column (f): Physics I
Column (g): Biology II
Column (h): Chemistry II
Column (i): Physics II

REFERENCES

Adams, D. (Ed). *Education in national development*. New York: David McKay Company, 1971.

Bross, I. D. J. *Designs for decision*. New York: The Macmillan Company, 1953.

Fisher, I. *The making of index numbers*. Boston: Houghton Mifflin Company, 1927.

Henle, S. J. *Systems for measuring and reporting the resources and activities of colleges and universities*. NSF Publication 67-15. Washington D.C.: U.S. Government Printing Office, 1967.

Kidd, J. L. *Texas teacher supply and demand*. Report prepared at the Texas Education Agency, Austin, Texas, 1978.

Knezevich, S. J. (Ed). *Creating appraisal and accountability systems*. San Francisco: Jossey-Bass, Inc., 1973.

Kronkosky, P. C. *Applied state-level educational decision making: A proposed model.* Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1970.

Rains, K. J. R. *Development and analysis of a model to describe the state of science education at the college level in Texas*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1979.

Sengupta, J. K. *Economic analysis for educational planning*. Baltimore: The Johns Hopkins University Press, 1972.

CLASSROOM PROCESSES AND THEIR RELATIONSHIPS WITH SHORT TERM OUTCOMES IN ELEMENTARY SCHOOL SCIENCE

CHARLES W. ANDERSON

INSTITUTE FOR RESEARCH ON TEACHING
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN

INTRODUCTION

Research on teaching effectiveness in the "basic skills" subjects of reading and mathematics made important advances during the 1970's. These advances were due in part to a number of process-outcome studies (reviewed by Brophy, 1979), in which data on teacher behavior and other occurrences in classrooms were related to student learning outcomes, which were typically measured as gains on standardized achievement tests. These studies represented an important advance in teacher effectiveness research for a number of reasons. Among these reasons were their emphasis on observation of "natural" classroom situations and the development of observation systems that provided more detailed and context-specific data than earlier systems.

The major process-outcome studies were limited in a number of ways, however, including their almost exclusive emphasis on reading and mathematics. The study described here was designed to extend process-outcome research methods to the study of science teaching in elementary schools.

The procedures used in earlier process-outcome studies, though appropriate for reading and mathematics, cannot be applied directly to the analysis of elementary school science instruction. Therefore adaptations of those procedures were necessary, including modifications of both the process measures and the outcome measures. For example, some processes, such as the use of manipulative science materials, are rarely observed during basic skills instruction. Therefore, a new observation system was designed that described classroom processes that are unique or likely to be meaningful for science instruction.

Standardized achievement tests were used as outcome measures in earlier research, but their use in the study of science teaching was unsatisfactory for two reasons. First, there are no widely accepted tests of achievement in elementary school science. Second, tests of student learning do not address one of the most important problems in elementary school science teaching today: the reluctance of many elementary school teachers to teach science at all (Stake and Easely, 1978; Weiss, 1978). For these reasons short-term outcome measures of student behavior and teacher satisfaction were developed and used in this study. These short-term outcome measures had the additional advantage that they could be incorporated into the observation system and collected on a lesson-by-lesson basis.

Thus the study described below departed from traditional teacher effectiveness research in the choice of process and outcome measures, while maintaining its emphasis on detailed and context-specific observation of "natural" classroom teaching. Results shed light on the nature of elementary school science instruction as it is experienced by both teachers and students, and on the crucial roles that elementary school teachers play as they decide how (or whether) science is taught.

RESEARCH DESIGN

SETTING

The study was conducted in a small, suburban school district. Both teachers and students were almost entirely Anglo, and most of the students came from middle to upper middle-class families. Achievement on standardized tests was above national norms. Observational data were collected during the 1978-79 school year. All 22 teachers in the district in grades two through five who taught science were included in the study.

Six teachers in the second grade were using materials from the Science Curriculum Improvement Study (SCIS). The SCIS curriculum emphasizes the use of manipulative materials and "hands-on" experience, generally with class discussions to reinforce and structure what the students learn from their experiences. Books and audio-visual materials were rarely used.

Teachers in the third, fourth, and fifth grades used a variety of materials, including locally developed materials, textbooks, and audio-visual materials. Lessons taught by the seven third-grade and eight fourth-grade teachers included fewer experiences with manipulative materials than second-grade lessons. The fifth-grade teacher, who was the only science specialist among the observed teachers, used a wide variety of manipulative materials, books and audio-visual materials.

THE OBSERVATION SYSTEM

The observation system used in this study was based to some extent on systems used for other process-outcome studies (Brophy and Evertson, Note 1; Brophy, Evertson, Baum, Crawford, and Edgar, Note 2) and on a review of research on classroom processes in which other promising variables were identified. Preliminary versions of the observation system were tested and modified according to their appropriateness to the setting and the research questions.

In its final form, the observation system provided both quantitative and qualitative data of 57 science lessons. A complete report on all data is given in Anderson (1979). Data reported in this paper include the following:

1. *Classroom descriptions* were "snapshots" of the classroom taken at ten-minute intervals during the lessons. Every ten minutes the observer scanned the room and described the classroom organization or format, the instructional materials being used, the noise level of the classroom, the location and activity of the teacher, and the activity of each student.
2. *The time spent in each of 12 classroom organizational formats*, such as whole-class discussions, small groups working on the same task,

small groups working on different tasks, or transitions, was recorded for each lesson.

3. The *instructional materials* used during each lesson were characterized in a similar manner. Fourteen categories were used to describe materials in use, including books, worksheets, manipulative science materials, and no materials in use.
4. During *class discussions* the observer recorded information about the teachers' statements and questions and the activities of the students. Student on task activities included raising hands to volunteer, answering questions asked by the teacher, and making relevant comments about the topic of the discussion. Off task activities such as irrelevant comments, social talking, and misbehavior were also recorded.

Although all data reported in this paper were recorded by a single observer, a second observer was trained to assess the reliability of the system. The two observers independently recorded data for 11 lessons. Reliability was assessed independently for each variable in the observation system, using procedures described in the section on data analysis.

OUTCOME MEASURES

Three types of measures were used to evaluate the success of the observed lessons:

1. *Teacher ratings*. After each of the observed lessons the teacher rated the success of the lesson according to three criteria: the students' enjoyment of the lesson, the ease of classroom management during the lesson, and the teacher's overall satisfaction with the lesson.
2. *Student behavior*. The average proportion of students off task was calculated for each of the observed lessons. This was selected as an outcome measure because of the well-documented association between measures of student learning and student on task behavior in a variety of different school settings (Berliner, 1979; Brophy, 1979; Rosenshine, 1979).
3. *Observer ratings*. After each lesson the observer rated the success of each lesson according to three criteria: student learning, student enjoyment, and the efficiency with which the teacher managed the classroom. Observer ratings and teacher ratings were taken independently. The observer was an experienced elementary school science teacher who was familiar with the science program materials used by the teachers.

Procedures used to assess the reliability and validity of the outcome measures are described in detail in Anderson, Note 1. Inter-observer reliability was .74 for the proportion of students off task and .86, .90, and .87 for the three observer ratings. The possibility of "observer drift" in the ratings was also checked by having the observer re-evaluate five lessons in the basis of the narrative notes and other information from the original observation forms. The elapsed time between the first and the second ratings was from three to six months, and the second ratings were done "blind" to the first. Fourteen of the 15 pairs of ratings agreed within one point on a nine-point scale.

DATA ANALYSIS

Since the short-term outcome measures were collected for each lesson, it was possible to use the lesson, rather than the teacher, as the unit of analysis for this study. Thus, the analyses conducted for this study produced comparisons of successful and unsuccessful *lessons*, comparisons of successful and unsuccessful *teachers*.

The data from the observation system were used to generate 234 classroom behavior variables, each of which described some small portion of the events during a given lesson. All of the classroom behavior variables in combination provided a detailed characterization of the lesson.

Two types of relationships between the classroom behavior variables and the outcome measures were examined:

1. *Simple correlations* were calculated between each of the classroom behavior variables and two types of outcome measures (student behavior and observer ratings). These correlations address the question of how more successful lessons were different from the less successful lessons without regard to characteristics of individual teachers.
2. *Linear regression equations* were used to examine the relationship between within-teacher variation for each classroom behavior variable and two types of outcome measures (student behavior and teacher ratings). This method of analysis addressed the question of how more successful lessons differed from less successful lessons for a given teacher.

Reliability was assessed by calculating inter-observer correlations for each of the 234 classroom behavior variables. Variables whose reliability coefficients were not significant at the .05 level (equivalent to $r=.55$ for 11 lessons) are indicated in the tables. Most of the variables that failed to achieve this criterion described very low-frequency events which were not observed often enough to establish reliability. The use of descriptive data from such variables remains appropriate (i.e., the observers agreed that those events rarely occurred).

RESULTS

DESCRIPTIVE DATA

These results are based on the observation of 57 science lessons taught to second-grade students, 22 to third-grade students, 18 to fourth-grade students, and 3 to fifth-grade students. The average number of students present was 22.7. The lessons ranged in length from 23 to 86 minutes. The average lesson was 39.5 minutes long.

Science was an "afternoon subject" for most of the observed teachers.

Although the school day started at 8:00 a.m. for most classes, only two of the observed lessons began before 10:00 a.m., while 43 began after 1:00 p.m.

During each of the observed lessons, the observer completed a classroom description every 10 minutes for a total of 212 classroom descriptions. Each classroom description included observation and classification of the behavior of every student present. Summary statistics describing observed student behavior are presented in Table 1.

The students spent most of their time listening to the teacher or to other students. Students also spent a considerable amount of time in off task activities such as "doing nothing" and social talk. This is not an indication that the observed classes were chaotic (they generally were not) or that students often failed to complete assigned work. Rather, most classes usually had a few students off task, and there were periods during most classes when many students had nothing to do. They waited to sharpen pencils; they waited for the teacher to answer their questions; they waited for materials to be distributed or picked up; they waited after they had finished an activity for a new activity to begin.

Table 1

HOW OBSERVED STUDENTS SPENT THEIR TIME IN SCIENCE CLASSES

Activity	Proportion of total time (%)
On task behavior	
Listening	.58
Writing	.09
Using science materials	.06
Talking about lesson-related matters	.03
Procedural tasks	.01
Reading assigned material[a]	.01
Off task behavior	
"Doing Nothing" (Off task, but not disturbing others)	.13
Social talk	.08
Misbehaving	.003

[a] Variables with reliability coefficient not significant at .05 levels.

Students used manipulative science materials in 25% of the observed lessons, but only 6% of the observed student time was spent in using manipulative science materials. This is a reflection of the fact that even in "hands-on" lessons the students spent most of their time doing other things, such as listening to instructions, discussing their results, waiting for materials to be distributed, or cleaning up.

The first column of figures in Table 2 indicates the proportion of time allocated by the teachers to a variety of classroom organizational formats; the second column indicates the average proportion of students observed off task when classrooms were in that format.

Very few lessons were observed in which the format stayed the same for the entire period. A typical materials-oriented lesson, for instance, included: (a) whole class discussions of the problem for the day, (b) procedural directions regarding use of materials, (c) transitions in which materials were distributed and cleaned up, and (d) individual or small group work with materials.

Transitions were unavoidable in most of the observed lessons — the alternative was the monotony of doing the same thing for an entire period. Transitions were often a source of management problems, however. On the average, 62% of the students were off task during the observed transitions.

Table 2

CLASSROOM ORGANIZATIONAL FORMATS

Format	Proportion of time allocated	Proportion of students off task
Whole class academic activities	.64	.14
Transitions	.09	.62
Students working individually on the same task	.08	.25
Whole class using audio visual materials	.07	.20
Students working in small groups on the same task	.03	.32
Procedural directions	.03	.18
All others[a]	.03	.20-1.00

[a] Student presentations, students working individually on tests, students working individually on different tasks, small groups working on different tasks, dead time, and "other."

The teachers tended to be most successful in avoiding off task student behavior during whole class discussions. The teachers also allocated almost two-thirds of the observed class time to this format. Whole class academic activities in the observed classes could almost always be described as discussions or recitations rather than as lectures. The teachers rarely talked for sustained periods without interruption, and rates of student activity were generally high. The rates at which a variety of student activities were observed are summarized in Table 3.

The rates at which teachers and students engaged in various types of activities were calculated by dividing the time spent in class discussions into three-minute segments. Thus, all of the data presented in this section are rates per three minutes of class discussion. Only 42% of the three-minute segments during the observed classes contained as much as 30 seconds of continuous teacher talk. Teacher questions were much more common; 83% of the observed time segments contained at least one teacher question.

Table 3

STUDENT ACTIVITIES DURING CLASS DISCUSSION

Activity	Number of students
Volunteered by raising a hand	7.2
Called out an answer, question, or comment[a]	2.3
Responded to a teacher's question as a volunteer	3.3
Participated in a choral response	6.2
Participated in a voting response	4.7
Responded to a teacher's question as a nonvolunteer	.2
Responded to a teacher's question in writing[a]	0.0
Made a relevant question or comment	1.2
Made an irrelevant question or comment	.6
Talked continuously for 10 seconds or more	.8
Talked socially	3.3
Misbehaved[a]	.1

Note: Average number of students who engaged in each activity during a 3-minute time segment are listed.

[a] Variables with reliability coefficient not significant at .05 level.

It is apparent from Table 3 that students were active during discussions in a variety of ways. Some of those activities, such as volunteering to answer questions or making relevant comments, were generally sanctioned by the teachers. Others, such as social talk and misbehavior, were generally regarded with disfavor. Student callouts and irrelevant questions or comments met with varied reactions from the observed teachers, who sometimes disapproved and sometimes treated them as legitimate contributions to the discussion.

The preponderance of whole class discussions was apparently subject matter specific for the observed teachers. Several of them indicated in informal conversations that they depended extensively on other forms of classroom organization, such as reading groups or individual seatwork, while teaching other subjects. Research on teachers' choices of classroom organization in other subjects (Evertson, Anderson, Anderson, and Brophy, 1979; Good and Beckerman, 1978; Rosenshine, 1980) has generally found a higher incidence of individual seatwork and small-group activities than reported here.

The instructional materials that were used in the observed lessons and associated levels of student off task behavior are summarized in Table 4.

Table 4

TEACHERS' CHOICES OF MATERIALS FOR TEACHING SCIENCE LESSONS

Type of materials	Proportion of time allocated	Proportion of students off task
No materials	.34	.20
Manipulative science materials	.12	.22
Books	.12	.12
Worksheets not made by teacher	.11	.20
Mixed non-manipulative materials	.08	.20
Students' own paper or notebooks	.04	.19
All others[a]	.12	.04-.64

[a] Mixed materials including manipulative science materials, records or audio tapes, tests, workbooks, teacher-prepared worksheets, art materials, games or puzzles, "other."

Two facts are apparent from Table 4. First, there was considerably more variety in teachers' choices of materials than of classroom organizational formats. Second, choices of materials were less strongly associated with the proportion of students off task than choices of classroom organization. The two patterns are probably connected; the teachers were more willing to use a variety of materials because their choices of materials did not markedly affect their control over student behavior.

RELATIONSHIPS INVOLVING OUTCOME MEASURES

Relationships between selected variables and the short-term outcome measures used in this study are as follows:

A. Relationships among outcome measures:

Analysis of results revealed that there was a general tendency for all of the outcome measures to be positively associated (except for the proportion of students off task, which is negatively related to all other outcome measures). In other words, lessons that were successful according to one criterion tended also to be successful according to the other criteria, and the judgments of the observer and the teachers generally agreed.

Within this pattern of overall consistency, the apparent importance that the teachers placed on classroom management is conspicuous. The observer's rating of overall management efficiency, for instance, is associated more strongly than other observer ratings with teacher satisfaction. The importance that the teachers placed on management is apparent in other results as well. Evidence presented above, for instance, suggested that teachers' decisions concerning materials and classroom organization were affected by concerns about keeping students on task. The available evidence indicates that these concerns are not inappropriate; management success has consistently been associated with teaching effectiveness in basic-skills instruction (Brophy, 1979; Rosenshine, 1979).

B. Variables describing student behavior:

Analysis of results revealed no relationship between these variables and teacher ratings. The observer, but not the teachers, tended to be less satisfied with lessons in which large numbers of students were observed off task, but not disturbing others. The teachers indicated dissatisfaction only when off task behavior was accompanied by social talk. Three of the nine variables describing student behavior (see Table 1 for a complete list) were associated with the outcome measures at a greater-than-chance level.

C. Variables describing classroom organizational format and materials:

The most notable result here is the general absence of significant relationships. None of the commonly observed formats (see Table 2 for a complete list) and only one type of commonly observed materials (see Table 4) was significantly associated with outcome measures.

A possible explanation for the paucity of significant relationships is that teachers tend to make instructional decisions that maximize their control over student behavior and their own satisfaction. Thus, they simply avoid materials and classroom formats that create problems for them. The one exception to the pattern of non-significant relationships is consistent with this interpretation. The curriculum programs that some teachers were following called for the use of manipulative science materials, but the teachers tended to use them reluctantly and to be less satisfied with the lessons in which they were used.

D. Variables describing teacher and student behavior during class discussions:

There were a large number of significant relationships between these variables and outcome measures. Teachers had less control over student behavior during class discussions than over their own choices of classroom organization and materials; thus, events during class discussions were more likely to affect their satisfaction with the lesson.

Both the teachers and the observer preferred lessons in which many students participated actively while "staying within the rules." Predictably, irrelevant questions or comments, social talk, and other types of off task behavior were regarded negatively by both the teachers and the observer.

Both the teachers and the observer tended to give higher ratings to lessons in which rates of individual student responses were high. The teachers often made use of "group" response modes, such as choral responses or voting responses (see Table 3). Although these response modes increased the rate of student participation, none of the variables describing "group" responses was significantly associated with the outcome measures.

The use of process questions (questions that require explanations from students rather than single-word or single-phrase answers) was positively associated with both teacher and observer ratings. The pattern of association with the outcome measures indicates that the teachers often would have liked to have asked more process questions than they did, but that they may have failed to create a context in which process questions were appropriate. Teachers could ask few process questions, for instance, when many students were not paying attention, or when students did not understand basic facts about the topic under discussion. It may be that failure to consider the problems involved in creating an appropriate context accounts for some of the confusing and contradictory results of the research on the efficacy of "higher-cognitive-level" questions (Winne, 1979; McGlathery, 1978; Rosenshine, 1977).

DISCUSSION

The results of this study delineate a challenge that must be faced by both elementary school teachers and other educators. The teachers who participated in this study depended heavily on whole class discussions as a means of teaching science, with occasional breaks for the use of manipulative science materials or audio-visual materials. When they did use manipulative science materials, they were often dissatisfied with the lessons that they taught.

Some understandable concerns probably motivated this pattern of teaching. The teachers were generally unsure of their knowledge of science, and they needed to maintain order in their classrooms. Thus, they opted for organizational patterns and student activities that were easy to manage and that allowed the teacher to control the information presented and discussed during the lesson. For these teachers, manipulative science materials were associated with uncertainty and loss of control; they perceived their teaching as more successful (and had few students off task) when manipulative science materials were not involved.

However, there is reason to be concerned about the quality of learning that may have occurred as a result of these instructional patterns. In combination, reading, writing, using manipulative materials, and talking about science occupied only 19% of the students' classroom time; students were either listening or off task 81% of the time. The generally passive learning style implied by these figures is quite different from that generally endorsed by science educators, who have long argued that activity-based science instruction offers superior opportunities for student learning (Barufaldi and Swift, 1977; Rowe, 1978).

Many teachers experience the tension between the practices advocated by science educators, which are implicit in most teacher training and science program materials, and their own needs for simplicity and control. All too often, the result is avoidance of science, as documented by Weiss (1978) and Stake and Easeley (1978). The teachers who participated in this study did not avoid science, but neither their training nor the science program materials available enabled them to provide the kind of instruction advocated by science educators. In science as in other subjects in the elementary school curriculum, the challenge of designing instructional techniques that meet the needs of students, maintain order in the classroom group, and recognize the human limitations of teachers is an exceedingly difficult one.

NOTES

1. Brophy, E., & Evertson, M. *Texas teacher effectiveness study: Classroom coding manual.* The Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, 1976.
2. Brophy, J., Evertson, C., Baum, M., Crawford, J., & Edgar D. Junior high school study: Coding manual (R&D Report No. 4014). The Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, 1975.

REFERENCES

Anderson, C. W. *An observational study of classroom management and information structuring in elementary school science lessons.* Unpublished doctoral dissertation, the University of Texas at Austin, Austin, Texas, 1979.

Barufaldi, J. P., & Swift, J. W. Children learning to read should experience science. *The Reading Teacher*, January, 1977, 388-393.

Berliner, D. C. Tempus educare. In P. L. Peterson and H. J. Walberg, (Eds.), *Research on teaching*. Berkeley: McCutchan, 1979.

Brophy, J. E. Advances in teacher effectiveness research. *Journal of Classroom Interaction*, 1979, *15*, 1-7.

Brophy, J. E. & Evertson, C. M. *Learning from teaching: A developmental perspective.* Boston: Allyn and Bacon, Inc., 1976.

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American Educational Research Journal*, 1980, *17* (1) 43-60.

Good, T. L., & Beckerman, T. M. Time on task: A naturalistic study in sixth-grade classrooms. *The Elementary School Journal*, 1978, *78* (3), 193-201.

McGlatherey, G. Analyzing the questioning behaviors of science teachers. In Mary Budd Rowe (Ed.), *What Research Says to the Science Teacher (Volume 1).* Washington, D.C.: National Science Teachers Association, 1978.

Rosenshine, B. V. Content, time, and direct instruction, P. L. Peterson and Herbert J. Walberg, (Eds.), *Research on teaching*, Berkeley: McCutchan, 1979.

Rosenshine, B. V. How time is spent in elementary school classrooms. In Carolyn Denham and Ann Liebermen, (Eds.), *Time to learn.* National Institute of Education, May, 1980.

Rowe, M. B. *Teaching science as continuous inquiry: A basic* (2nd ed.). New York: McGraw Hill, 1978.

Stake, R. E., & Easley, J. *Case studies in science education, Volume II: Design, overview, and general findings.* Center for Instructional Research and Curriculum Evaluation, University of Illinois, 1978.

Weiss, I. R. *Report of the 1977 national survey of science, mathematics, and social studies education.* Center for Education Research and Evaluation, Research Triangle Park, North Carolina, 1978.

Winne, P. H. Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research*, 1979, *49* (1), 13-50.

EXPECTATIONS OF THE SCIENTIFIC ENTERPRISE

SUSAN HELMS
DEPARTMENT OF CHEMISTRY AND PHYSICS
HARDIN-SIMMONS UNIVERSITY
ABILENE, TEXAS

INTRODUCTION

Science educators' concern for the development of scientific literacy in the non-science major has led to the emergence of a variety of science courses designed especially for the non-science major. Evaluation of these courses and their effectiveness in developing various components of scientific literacy is a current interest.

Considerable concern has recently been expressed about the nature and extent of the expectations of the scientific enterprise held by members of the general public (e.g., Abelson, 1970; Harrison, 1974; Shils, 1974; and Scott, 1978). There are three major contexts in which realistic expectations of the scientific enterprise are particularly important. First, the educated citizen must make decisions based on a realistic understanding of what to expect from the scientific enterprise. Second, in a democratic society much of the support for the continued functioning of the scientific enterprise depends on society's expectations. Research shows that inflated expectations, as well as negative or low expectations, produce lowered evaluations in the public's mind (Anderson, 1973; Oliver, 1977). Third, realistic expectations concerning one of the primary sources of change in a rapidly changing society can provide a strong basis for coping with that change (Hurd, 1970).

A desirable component of scientific literacy, then, would be an understanding of the nature of the scientific enterprise and its relationships to the rest of society sufficient to produce realistic expectations of the scientific enterprise. Realistic expectations of the scientific enterprise can thus be seen as more than just a component of scientific literacy, for their development requires the application, or transfer, of many components of scientific literacy to a consideration of the future.

PROBLEM

The major problem investigated by this study was: Are college non-science students' expectations of the scientific enterprise related to the students' educational experiences?

Important subproblems were:

I. Is there any relationship between non-academic background factors (gender, fiction and non-fiction science reading, and science television viewing) and college non-science majors expectations of the scientific enterprise?

II. Is there any relationship between background factors (classification, previous science background in high school and college, and percent of previous science course time spent on various course content emphases) and college non-science majors expectations of the scientific enterprise?

III. Is there any relationship between specific aspects of science courses for non-science majors (content emphases, teaching methodologies, and whether or not laboratory work is included) and students' expectations of scientific enterprise?

PURPOSE

The purpose of the study was to determine what kinds of learning experiences may be related to the development of realistic expectations of the scientific enterprise in college non-science majors. Specifically, the study examined what kinds of academic and non-academic background factors influence the development of realistic expectations, and what aspects of students' current science courses may be related to realistic expectations.

The development of realistic expectations of the scientific enterprise was treated as an application of various components of scientific literacy to a consideration of future events. The problem thus becomes one of transfer of learning. Though transfer of learning studies usually deal with much simpler contexts, several principles of transfer do appear relevant:

1. The student must understand a concept for it to be available for transfer (Royer, 1978).
2. When dealing with abstract concepts, analogies, concrete referents, or cognitive structures are useful in facilitating the understanding necessary to subsequent transfer (Reynolds, 1966; Royer & Cable, 1976).
3. Practice in transfer facilitates subsequent transfer (Bourne & Guy, 1968).
4. The student must recognize opportunities for transfer (Bigge, 1976).

Using the above described principles of transfer of learning, course components and academic and non-academic background factors which might be expected to contribute to the development of realistic expectations of the scientific enterprise were identified.

RESEARCH DESIGN

SAMPLE

For practical reasons, the study was limited to students enrolled in courses for non-science majors in the area of chemistry. The sample was comprised of 1,014 students enrolled in 24 chemistry courses for non-science students at 19 colleges and universities. The Expectations of the Scientific Enterprise instrument along with a 10-item background questionnaire was administered to the students in their chemistry courses late in the spring term, 1979. Additionally, instructors for each course completed a survey dealing with the content emphases, teaching methodologies, and laboratory work involved in their courses.

INSTRUMENTATION

The realistic-unrealistic dimension of students' expectations of the scientific enterprise was assessed using a 30-item, multiple response, Likert-type instrument. Students were asked to respond with one of five choices ranging from "strongly agree" to "strongly disagree" to expectation statements like the following:

In the future I expect that:

> with sufficient research, scientists will develop inexpensive, pollution-free, inexhaustible energy sources.
>
> continued scientific research will probably lead to changes in our way of life.
>
> the results of scientific research will eventually eliminate mankind's problems.

The content validity of the Expectations of the Scientific Enterprise (ESE) instrument was established by panel review of the items. Construct validity was determined by the method of contrasted groups. Reliability of the instrument was evaluated both for internal consistency (a Cronbach alpha of 0.79 for $N=1014$), and for stability (a test-retest correlation coefficient of 0.65 for $N=49$).

DATA ANALYSIS

Responses were scored on the ESE by giving five points for a "strongly agree" response on a realistic item or a "strongly disagree" response on an unrealistic item; four points were given for an "agree" response on a realistic or "disagree" response on an unrealistic item; and so on, with a score of 1 being the lowest possible item score. The ESE instrument score was simply the sum of the thirty item scores.

A Standard Coding Form was used as an answer sheet for the questionnaire and ESE responses. The answer sheets were read by optical scanner and responses transferred to punched cards for computer processing. Instructors' responses concerning course characteristics were also transferred to punched data cards.

HYPOTHESES

The following hypotheses, stated in the null form, were developed to determine what sorts of relationships, if any, exist between the formation of realistic expectations of the scientific enterprise and students' background and learning experiences:

I. There will be no statistically significant* correlation between students' ESE scores and:
   a. the students' genders.
   b. the frequency of:

*the 0.05 level was used in establishing statistical significance throughout the study.

        (1) nonfiction science television viewing.
        (2) nonficton science reading.
        (3) science fiction reading.

II. There will be no statistically significant correlation between students' ESE scores and:
  a. students' academic classifications.
  b. students' cumulative high school science grade points (the product of the grade point average and number of courses).
  c. students' cumulative college science grade points.
  d. the percent of course time in previous science courses spent:
    (1) learning facts and principles of an area of science.
    (2) studying the history of science (or of an area of science).
    (3) learning about scientific methodology and techniques.
    (4) learning about how science applies to practical problems and what the role of science is in society.

III. a. There will be no statistically significant correlation between students' ESE scores and the percent of course time in the current science course spent:
    (1) developing the basic principles of an area of science.
    (2) studying scientific methodology.
    (3) performing numerical calculations.
    (4) presenting an historical development of the area of science.
    (5) studying specific applications of the area of science to practical problems.
    (6) studying specific applications of the area of science to societal problems.
  b. There will be no statistically significant correlation between students' ESE scores and the percent of class time spent in:
    (1) lecture.
    (2) demonstration.
    (3) discussion.
  c. There will be no statistically significant difference between the mean scores on the ESE instrument of students in courses involving laboratory work and those not involving laboratory work.

RESULTS

Correlation coefficients of non-academic background factors with ESE total scores are given in Table 1. There was no statistically significant correlation between students' genders and ESE scores, so hypothesis I(a) could not be rejected. All three factors relating to non-academic sources of information concerning the science and science-society relationships produced statistically significant, though small, correlations with ESE scores. Both frequency of science nonfiction television viewing and nonfiction science reading correlated positively with ESE scores, while frequency of science fiction reading correlated negatively with ESE scores. Thus hypotheses I b, c, and d were rejected.

Table 1

CORRELATION OF NON-ACADEMIC BACKGROUND FACTORS WITH ESE TOTAL SCORES

Background Factor	Correlation Coefficient	Level of Significance	N of Cases
Gender	-0.0258	0.206	1014
Frequency of Viewing T.V. Science	0.0536	0.044	1010
Frequency of Reading Science Non-Fiction	0.1542	0.001	1013
Frequence of Reading Science Fiction	-0.0760	0.008	1014

Correlation coefficients of academic background factors with ESE total scores are given in Table 2. Students' classifications, high school science background (measured as cumulative science grade points — the product of grade points and the number of courses), and college science background (again as cumulative science grade points) produced positive and statistically significant, though small, correlations with ESE total scores. Therefore hypotheses II a, b, and c were rejected.

In considering specific content emphasis of previous college science courses, the percent course time spent studying the history of science, learning about scientific methodology and technique, and learning about practical applications of science and the role of science in society all produced small, negative, statistically significant correlations, causing hypotheses II d (2), (3), and (4) to be rejected. Hypothesis II d (1) could not be rejected as no significant correlation was found between ESE scores and percent course time spent learning facts and principles of an area of science.

Table 2

CORRELATION COEFFICIENTS FOR ACADEMIC BACKGROUND FACTORS WITH ESE TOTAL SCORES

Background Factor	Correlation Coefficient	Level of Significance	N of Cases
Classification	0.0671	0.017	1007
High School Science	0.1668	0.001	1014
College Science	0.1362	0.001	1014
Previous College Science Content:			
Facts and Principles	0.0073	0.419	801
History	-0.1478	0.001	807
Methodology	-0.1090	0.001	806
Practical Applications & Science-Society	-0.0793	0.012	808

Since nine background factors were found to be significantly correlated with ESE scores at the 0.05 level, all correlation coefficients for current course components with ESE scores were calculated as partial correlation coefficients, controlled for the effects of these statistically significant background factors.

Table 3 includes the partial correlation coefficients for current course content emphases and ESE total scores. Only one of the course content emphases, percent class time spent studying the basic principles of an area of science, correlated significantly with ESE scores. Thus hypothesis III a (1) was rejected, and hypotheses III a (2) - (6) were not rejected.

Table 3

CORRELATION AND PARTIAL CORRELATION COEFFICIENTS FOR COURSE CONTENT EMPHASIS AND ESE TOTAL SCORES

Course Content Emphasis	Correlations Coefficient (Sig. Level)	Partial Correlation Coefficient (Sig. Level)
Basic Principles	-0.0046 (0.491)	0.6987 (0.002)
Scientific Methodology	-0.1671 (0.218)	-0.2881 (0.151)
Numerical Calculations	-0.0082 (0.485)	0.0545 (0.423)
History	-0.2317 (0.138)	-0.0538 (0.425)
Practical Applications	0.3203 (0.064)	-0.2955 (0.142)
Societal Problems	-0.0275 (0.449)	-0.2655 (0.170)

Partial correlation coefficients between instructional methodologies and ESE scores are shown in Table 4. None of the coefficients was significant at the 0.05 level, therefore hypotheses III b (1) - (3) were not rejected.

Of the 24 courses in study, 15 had laboratory work as a regular course component, and nine had no laboratory work. Neither a simple t-test nor analysis of covariance (using the five statistically significant background factors with the largest ESE correlations as covariates) showed any statistically significant difference between courses with and without laboratory work. Hypothesis III c was therefore not rejected.

Table 4

CORRELATION AND PARTIAL CORRELATION COEFFICIENTS FOR COURSE TEACHING METHODOLOGIES AND TOTAL ESE SCORES

Percent of Class Time Spent in:	Correlation Coefficient (Sig. Level)	Partial Correlation Coefficient (Sig. Level)
Lecture	0.3094 (0.071)	-0.2520 (0.182)
Demonstration	-0.0649 (0.382)	0.4258 (0.057)
Discussion	0.1125 (0.300)	-0.3977 (0.094)

DISCUSSION

The relationships shown between ESE scores and science television viewing and reading about science (both fiction and nonfiction) suggest that exposure to non-academic sources of information about science and science-society relationships may have some effect on the development of realistic expectations of the scientific enterprise. That academic exposure to science may also contribute to the formation of realistic expectations is shown by the direct relationship between both high school and prior college science background and ESE scores.

Prior science courses content emphasis on history of science, scientific methodology, and practical applications of science in society were found to be negatively related to realistic expectations while current course emphasis on basic principles was positively related to realistic expectations.

For this group of non-science majors, the first component necessary for the transfer of learning, an understanding of the concept to be transferred, may be lacking. Students who do not possess basic knowledge and fundamental concepts in science may not be helped by efforts to put scientific information into a practical or societal framework, and time so spent may detract from time available in a course for development of the fundamentals they may be lacking.

Since these conclusions are based on a correlational study, no cause-and-effect relationships have been established. If future research could identify cause-and-effect relationships consistent with the correlations found in this exploratory study, then the implications could be of importance to those involved in the development and improvement of college science courses for non-science majors. Specifically, there may be a need to place emphasis on the development of basic scientific principles and concepts in order to promote the formation of realistic expectations of the scientific enterprise.

REFERENCES

Abelson, P. H. Science and immediate social goals. *Science, 169,* 1970, 721.

Anderson, R. E. Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of Marketing Research, 10,* 1973, 38-44.

Bigge, M. L. *Learning theories for teachers.* New York: Harper and Row, 1976.

Bourne, L. E. & Guy, D. E. Learning conceptual rules: I. Some inter-rule transfer effects. *Journal of Experimental Psychology, 76,* 1968, 423-429.

Harrison, A. J. Provocative opinion — Chemical education and the expectations of society. *Journal of Chemical Education, 51,* (9), 1974, 569.

Hurd, P. D. Scientific enlightenment for an age of science. *The Science Teacher, 37,* (1), 1970, 13-15.

Oliver, R. L. Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology, 62,* 1977, 480-486.

Reynolds, J. H. Cognitive Transfer in Verbal Learning. *Journal of Educational Psychology, 57,* 1966, 382-388.

Royer, J. M. Theories of learning transfer. Technical Report No. 79, Center for the Study of Reading, University of Illinois at Urbana-Champaign, 1978 (ED-149-326).

Royer, J. M., & Cable, G. W. Illustrations, analogies, and facilitative transfer in prose learning. *Journal of Educational Psychology, 68,* 1976, 205-209.

Scott, F. Science as magic. *American Laboratory, 10,* (9), 1978, 6.

Shils, E. Faith, Utility, and the legitimacy of science. *Daedalus, 103,* 1974 1-15.

A SCIENCE LABORATORY SAFETY COMPUTER PROGRAM SIMULATION FOR PRESERVICE TEACHERS

FLOYD D. PLOEGER
SOUTHWEST TEXAS STATE UNIVERSITY
SAN MARCOS, TEXAS

Many science educators advocate the use of science classroom laboratory activities. However, the teacher is considered liable for the consequences of accidents that occur in the classroom (Scott and Hazari, 1978). Safety is a component of the science classroom laboratory and a concern of science teachers (National Science Teachers Association, 1978). Safety in the science classroom laboratory can have both immediate and far reaching consequences. Students and teachers in the classroom may be immediately affected as a direct result of a particular incident. The behavior patterns of students established in the school science classroom laboratory may well decide the relative safety of the individual and the class as a whole. Future coworkers of students in job related environments may suffer the consequences or reap the benefits of safety procedures learned in the science classroom laboratory.

Providing for the safety of the classroom is a process of recognition (Texas Education Agency, 1971). The teacher must be able to demonstrate the capability of correctly responding to a set of circumstances in order to avoid an accident (American Chemical Society, 1974). Some situations involving safety allow more lead time before action must be taken than do other situations. In the science classroom laboratory, several situations involving safety may be occurring within a particular time frame. The science classroom teacher must be able to ascertain which situations require the more immediate action and which situations lend themselves to deferred action.

Many of the conditions which may be encountered in the real situation or environment cannot be duplicated repetitively due to excessive cost, time constraints, danger or large space requirements (Jamison, Suppes, & Wells, 1974). For example, oil well crews must know what to do in order to prevent a "blowout" from occurring. Lives may be placed in danger and costs, both real money and environmental, may be astronomical. Computer program simulations permit the application of a number of useful features. The use of a computer program simulation provides for the safe reproduction of an environment or set of conditions at a considerably reduced cost.

PROBLEM

The problem under investigation was to determine the effect of the science laboratory safety computer program simulation on preservice teachers' classroom safety behavior. The following research questions were investigated:

1. Is there a pretest effect on the posttest prioritization of the hierarchy by the subjects?

2. Do the subjects transfer the prioritization of the hierarchy to another science classroom laboratory situation?
3. To what extent can it be inferred that the subjects' prioritization of the safety situations is a result of the interaction with the computer program simulation?

RESEARCH DESIGN

SAMPLE

The subjects for this study were the preservice teachers enrolled in the science methods courses at a large university in south central Texas, during the Spring semester, 1981. The subjects were all either at the junior or senior level in the College of Education. A total of 52 subjects were used for the study.

EXPERIMENTAL DESIGN

The Randomized Solomon Four-Group design (Campbell & Stanley, 1970) was chosen as the experimental design for this study (Table 1).

Table 1

SOLOMON FOUR-GROUP DESIGN

	Pretest	**No Pretest**	
Treatment	n =13	n = 13	n = 26
No Treatment	n = 13	n = 13	n = 26
	n = 26	n = 26	

TREATMENT

Each subject in the pretest group was presented with a black and white line drawing. The line drawing depicts 14 elementary school level students in an ongoing elementary school science classroom laboratory. Each of the depicted students is identified in the drawing by a clearly lettered name on the drawing next to the student. Of the 14 depicted students, 6 were judged to require action by the teacher in order to prevent an accident from occurring and insure that a safe environment would be maintained in the science classroom laboratory (Ploeger, 1980). Each subject was asked to view the black and white line drawing. They were then asked to identify the student in the drawing which they believed requires the most immediate teacher action. That is, the subjects were instructed to select the student depicted as being in the most dangerous situation and requiring the most immediate teacher action. The subjects were requested to use safety as the sole criteria for determining which students were to be dealt with. The subjects were requested to write the name of each student requiring a teacher action in rank order of the student requiring the most immediate teacher action to least immediate teacher action.

The subjects in the no pretest group were given a black and white line drawing depicting an adult and an elementary level child working in a kitchen. The subjects were asked to list the items in the depicted kitchen which they might reasonably expect to find in an elementary school level science classroom laboratory.

The subjects in the treatment group were asked to interact with the computer program simulation concerning safety in the elementary level school science classroom laboratory.

The subjects in the no treatment group were asked to interact with a computer program simulation involving gaming and probability (Truax, 1980).

All subjects were given the posttest which consisted of the black and white line drawing used by the pretest group. The instructions given to the posttest subjects were the same as the instructions which were given to the pretest subjects.

Finally, all subjects were given a redrawn black and white line drawing of the elementary school level science classroom laboratory in which the age level of the students were judged to be that of secondary school students. Each subject was again given the same instructions as were given in the pretest experience. This is referred to as the transfer measure.

INSTRUMENTATION

The treatment instrument used in this study is an interactive computer program simulation. The computer program is designed to simulate the complexity of the interactions that a teacher might have with students during a science classroom laboratory. The computer program simulation is designed to keep track of the students in the computer program simulated science classroom laboratory. The computer program records the interactions and removes students from the list of students needing to be dealt with by the teacher as the subjects work with the simulation. If an interaction with one of the students is interrupted by an emergency in the classroom, the computer program records the fact that the student has not been completely dealt with. The subjects must resolve all potential safety hazards in the science classroom laboratory before the computer program simulation will permit the subject to finish. Parsing by an "intelligent" key word recognition program allows subjects to type and enter sentences to which the computer program responds. The subject is informed as to the appropriateness of the request with respect to safety in the simulated science classroom laboratory.

The interaction with computer program is initiated by the subject based upon the black and white line drawing depiction of the elementary school level science classroom. Each subject is required to type and enter the name of the student in the line drawing which is believed to require the most urgent teacher action. After choosing a student, the subject must choose from several alternative actions which might be expected to occur. For example, one of the situations depicted a girl with long, untied hair in the process of attempting to light an open flame. The computer program simulation displays the following on the screen.

"Which of the following will you do?
A. You need to tie your hair back.
B. Get out of this classroom. You know the rules.
C. Stop! You could catch you hair on fire.
D. You need to get your hair cut by tomorrow."

The subject types the letter of the action which is believed to be the most desirable action. The computer program simulation is designed to evaluate the relative danger of the situation. If no more dangerous situation exists, the subject is allowed to continue interacting with the student which they have chosen until the hazardous situation has been resolved. The subject must correctly identify each depicted safety hazard in the black and white line drawing. All safety hazards must be dealt with before the subject is allowed to end the computer program simulation.

If the subject chooses a safety situation which is not the most hazardous, the subject is allowed to interact with the computer program simulation for a length of time. After a brief time period, the interaction is interrupted and the subject is presented with the message "You have a classroom EMERGENCY. Which of the following will you do?" The subject is presented with the emergency situation and forced to choose an alternative action. After the science classroom laboratory emergency has been successfully resolved, the subject is allowed to continue dealing with the remaining classroom hazards. Any interaction which is interrupted by an emergency is considered to be pending and must also be resolved. That is, it is assumed that any interaction with a student which is interrupted has not reached a logical conclusion and must be completed successfully.

DATA ANALYSIS

The safety situations were assigned a rank value with the most urgent safety situation receiving a rank value of 14 and the least urgent safety situation a rank value of 9. Each of the situations which require no teacher action have a rank value of 0. A standard correlation coefficient was calculated for each subject's ranking on each of the measures of the pretest, posttest and transfer measure as compared to the ranking established by the judges (Ploeger, 1980). Each subject's correlation coefficient was treated as a score for that subject. The mean correlation coefficient was calculated for the rankings of each of the groups of subjects in the Randomized Solomon Four-Group design. A comparison between the mean score value for each of the groups was made by using a simple t-test (Table 2).

Table 2

RESULTS OF COMPARISON OF MEAN SCORES

	Pretest	Posttest	Transfer
Treatment	t = 6.31 p < .001 ---->	p < .001 t = 7.19	p < .001 t = 4.05
No Treatment	t = .4089 p = N.S. ---->		

	Posttest	Transfer
Pretest	t =.3124	t = .6251
No Pretest	p = N.S.	p = N.S.

RESULTS

The data suggests that there are significant differences among the groups in the study. The data supports each of the research questions.

Question 1: Is there a pretest effect on the posttest prioritization by the subjects? As is indicated by the t-test and corresponding confidence level (Table 2), no significant difference was found between the pretest and no pretest groups on either the posttest or the transfer measure. It is inferred that there is no significant effect on the posttest scores nor the transfer scores by the pretest experience.

Question 2: Do the subjects transfer the prioritizations of the hierarchy to another science classroom laboratory situation? No significant difference was found between subjects' scores on the posttest measure and the transfer measure. It is reasonable to suggest that the subjects transfer the ability to rank the safety hazards in differing situations.

Question 3: To what extent can it be inferred that the subjects' prioritization of the safety situations is a result of the interaction with the computer program simulation? Analysis of the data (Table 2) suggests that there is a significant difference between the treatment group subjects' rankings on the pretest measure and the posttest measure. However, no significant difference was found between the no treatment subjects' rankings on the pretest measure and the posttest measure. The measure of the ranking of the safety situations by the subjects after the treatment differs significantly from the measure before the treatment. The difference is inferred to be independent of any effect caused by the pretest experience.

LIMITATIONS

The study takes advantage of the power of the Randomized Solomon Four-Group Design to determine the effect of the pretest on the posttest scores. The placement of the transfer measure, in time, immediately after the posttest measure is believed to serve to insure that "other" factors or experiences do not interfere with the experience of the computer program simulation.

The interactions which the subjects have with the computer program simulation are written on the computer screen. Thus, the failure to effectively simulate the verbal interactions normally found in the science classroom laboratory is believed to be a limitation of the study. It remains to be determined the extent to which the subjects' interactions with the computer program simulation have an effect on teacher behavior in the actual science classroom laboratory.

The line drawing used in this study is drawn in black and white. The science classroom laboratory is visually dynamic and in color. It is believed to be a limitation that the visuals in this study are significantly different from that of the real classroom as to require more study.

The subjects in this study represent a relatively small sample from the population of preservice teachers. It is believed that the computer program simulation should be used in a variety of settings to determine the effectiveness both with a larger population and in a variety of settings.

DISCUSSION

The results of this study suggest that a computer program simulation is effective in enhancing the ability of preservice teachers to recognize and correctly handle hazardous situations in the science classroom laboratory. The results suggest that preservice teachers were relatively unable to identify and prioritize safety hazards in the science classroom laboratory prior to interaction with the computer program simulation. It is believed that by interacting with the computer program simulation, preservice teachers can learn to recognize safety hazards in the real classroom. Since the computer program simulation is relatively inexpensive and free from danger, it is believed to be a viable teaching tool.

The implications of the study suggest that computer program simulations are potentially effective teaching tools. Many interactions in the classroom may be simulated by a computer program and in so doing free the teacher to interact with students in more humanistic tasks. The computer potentially may be programmed to respond to the user without suffering the range of emotions that may detract from the learning experience. This is not to suggest that certain humanistic emotional interactions are not valuable. However, anger and boredom from repetitive interactions are believed to be potentially detracting from the learning experience.

In this context, the tactic used in this particular program simulation is believed to be particularly valuable. This computer program simulation is designed so as not to require that the subject be informed of being "right" or "wrong." This simulation simply attempts to recreate the interactions found in the classroom. Actions taken by the subjects result in reactions by the computer program which closely resemble the interactions found in the classroom. That is, this simulation does not inform the subject that failing to correctly deal with a student is incorrect. Rather, the simulation informs the subject of the classroom situation after an interaction has taken place. By failing to stop a student from eating candy, the simulation informs the teacher that a lawsuit has been initiated on behalf of the parents of the poisoned student. Failing to insure that a student holding glass tubing knows proper techniques for cutting the tubing results in the teacher being informed that the student in question has just cut his hand.

The computer program simulation is designed to simulate the types of interactions encountered in the classroom. Interactions in the classroom are determined by the reactions of students and the teacher to situations. The interactions in the classroom are not carried out as a result of or reaction to "That's correct," or "No! Try again." It is believed that the major significance of this study is that the subjects are not told "right" and "wrong." The subjects are simply informed of the situation resulting from any action or lack of action by a teacher. The subjects must determine the appropriate action based upon the existing conditions as perceived by the subject. This simulation has been demonstrated to be effective and the effectiveness is believed to be largely a function of lack of directly informing the subject of the "rightness" or "wrongness" of their action.

REFERENCES

American Chemical Society. *Safety in academic chemistry laboratories.* Commission on Chemical Safety, 1974.

Campbell, D. T. & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally & Company, 1970.

Jamison, D., Suppes, P. & Wells, S. The effectiveness of alternative instructional media: A survey. *Review of Educational Research, 44,* 1974, 1-67.

National Science Teachers Association. *Safety in the secondary science classroom.* Washington, D.C.: National Science Teachers Association, 1978.

Ploeger, F. D. *SAFE: A model for teaching and assessing competencies in the secondary school science classroom.* Unpublished masters thesis, The University of Texas at Austin, Austin, Texas, 1980.

Scott, R. & Hazari, A. Liability in the academic chemistry laboratory. *Journal of Chemical Education, 55* (Al96), 1978.

Texas Education Agency. *Planning a safe and effective learning environment for science.* Bulletin 718, 1971.

Truax, R. H. *An interactive computer program simulation concerning probability.* Unpublished program, The University of Texas at Austin, Austin, Texas, 1980.

TEACHING THE MOST BASIC SKILL - OUR SURVIVAL DEPENDS ON IT

MARIANNE BETKOUSKI
UNIVERSITY OF NORTH FLORIDA
JACKSONVILLE, FLORIDA 32216

INTRODUCTION

"Teach the basic skills!" State legislatures, the general public and, consequently, local boards of education are repeating these words and the impact is being felt in curricular plans, testing procedures, and the everyday operations of the classroom. What is a basic skill? It would seem that such a skill would be defined as one needed to *survive* in our current complex society. Certainly, proficiency in reading, writing, and mathematical computation are of paramount importance in a person's efforts to become educated and self-supporting. Therefore, these are, indeed, basic skills. However, if that same person lacks the ability to cope with the uncertainty of life in these often chaotic times, then all the aforementioned "basic" skills are inoperative. The most basic skill depends on a person's whole-life perspective and problem-solving ability. This latter skill can be effectively tied into a science framework in which the scientific method of problem solving becomes a basic for implementing real-life coping skills.

Indeed, science educators are at a moment of a decision in their finding a defense by which to keep science in the schools. Budget cuts and "basic skills" emphasis are eroding attention to science instruction, especially at the elementary and secondary education levels. Ironically, students need the study of science process and science content now more than ever in our quickly changing technology-based society. Ignorance in a situation often leads to inability to cope with the effects of that situation. Science educators must become more vocal in their efforts to proselytize their cause and assure that science teaching is not impaired in the name of paring down the curricula in schools. If we believe in what we are doing as science educators, then we must convince those who make the decisions that science and survival are synonymous. Those who define a problem have a decided advantage in solving it.

One method by which to connect science teaching to the real needs in people's lives is to develop a sound rationale for what should be taught in the schools. In this regard, Shane (1981) has listed six necessities to improve the climate for learning:

1) the establishment of personal security;
2) the use of standards to . . . guide the individual learner rather than to norm or rank him/her;
3) the expansion of the scope of schools into the realms of mental, emotional, and physical health;
4) an emphasis on reasoning and problem-solving skills rather than on producing the "unexpected" answer only;

5) the opportunity for participation in group decision-making as one application of the democratic process;
6) the assurance of a curriculum that is demanding, yet interesting and relevant to the needs of a high technology society.

Another connection lies in an effort started in the early part of this century. The periodic re-examination of the Cardinal Principles of Secondary Education (Saylor, Alexander & Lewis, 1981, p. 179) has brought these goals, first established in 1918, more in line with the needs of the whole person. Originally, the areas covered were health, command of fundamental processes, worthy home membership, vocation, civic education, worthy use of leisure time and ethical character. In 1938, emphasis was placed on self-realization, human relations, economic efficiency and civic responsibility. Finally, in 1977, the need for improving human relations skills was reemphasized and the need for computer language skills (a direct result of technological advances) and the need for anticipatory problem-solving skills were added.

Given that there is a defense for a broad-based curriculum, a key issue emerges: Are the "whole life" needs of our students actually being neglected? The 1981 Gallup poll of public attitudes toward education (Gallup, 1981) still identified discipline and drug use as the number one and number two problems in our schools. Drug use is often self-medication for depression (Newsweek, 1974), and depression can be defined as a feeling of helplessness and hopelessness. Although few studies have actually produced statistics on depression in children and adolescents, research has connected adolescent depression with low self-esteem (Battle, 1980). A five year study of middle school teachers (Johnson, 1979) confirmed the difficulties that teachers have in diagnosing and responding to early signs of depression in their students. Short attention span, quickness to anger, and abnormal fear of failure can easily be attributed to student lack of cooperation rather than to feelings of real helplessness. If a student's basic needs are not being met, he/she is apt to "escape", either by resorting to discipline problems or to drug-taking.

It appears that educational practitioners can place more energy designing programs where students needs are a major focus. "Coping curricula" in schools are a possible means by which to help students to rechannel their restlessness and discontent into actual coping mechanisms with a problem-solving orientation. "Basic skills" need not be neglected in this endeavor. By way of hypothesis, learning in the true sense of the word can only be enhanced in a setting where needs are identified *and* met.

SCIENCE EDUCATION EMPHASIS

Science educators are in a position to lead a pioneering effort in the establishment of a "coping curriculum" which begins to address Shane's list of necessities to improve learning climates. A prototype plan in this area was suggested by the Texas Science Framework Revision Committee (Texas Education Agency, Science Section, 1974) when it proposed a matrix (Figure 1) to guide the development of minicourses in Texas secondary schools. The matrix integrates "What People Do" and "What Schools Do" and allows minicourse areas to emerge as the matrix cells. While goals identified in the matrix appear to be relevant to student lives, pilot projects designed to test the effectiveness of possible minicourses need to be constructed. The remainder of this paper addresses the design and evaluation of a minicourse which fits into the realm of column A of the matrix.

MINICOURSE RATIONALE

The minicourse was designed by the author as a two-week exposure to the physiological, biochemical, and psychological correlates of emotional stress. It was hoped that a student's life-stress coping skills could be enhanced by such a course, and that the student would feel in better control of his/her own fate as a result of the course experiences. Biofeedback researchers such as Brown (1974) hold that the physiologic and subjective body processes can be regulated by the individual. She professes that this regulation of body processes is enhanced if the individual has a certain amount of scientific information about the processes. Apparently, specific kinds of stress (e.g., test anxiety) can be alleviated through the application of the principle of "informational learning." Richardson defines "informational learning" as the ability of an individual to change his/her behavior in a desired direction if he/she is given accurate and useful information about the problem with which he/she is trying to cope, and he has gathered evidence to support this principle.

MINICOURSE DESIGN

The process by which the course was formulated followed the approach of Klopfer (1971), including designation of course objectives, selection of appropriate activities in each objective, and formative and summative evaluations. A teacher's guide was constructed, also, with detailed directions concerning laboratory set-ups and interpretation procedures. The overall course objective is as follows: The students will analyze certain psychological, physiological and biochemical aspects of emotional stress in order to understand what happens to their bodies in a state of stress and to take appropriate action to solve stress-producing problems.

Sixteen terminal behavioral objectives are included in the course which is presented in a 62 page booklet entitled, *Emotional stress: To fight or run away.* Activities range from the students' listing of their emotional and physiological reactions, listing their own stresses, observing other organisms (e.g., mealworms, Daphnia) react to various stimuli, and then "comparing notes." Students even build a simple electrochemical cell and compare it to a nerve cell. A major course emphasis is the biochemical individuality of each human being.

MINICOURSE EVALUATION

The minicourse was pilot tested with a group of ten teenage students in a central Texas city. As a result of the pilot study, course revisions such as clearer instructions in the booklet were implemented. The major study of course effectiveness involved the participation of six teachers and 431 secondary school biology students in a central Texas school district. Two treatments and a control group of approximately equal size were involved in this study. Treatment A consisted of the minicourse followed by a four day session on coping with test anxiety.

The "coping" segment made use of Richardson's (1973) self-paced manual of instructions geared to reducing test-taking anxiety. Treatment B consisted of the test anxiety coping treatment alone. The treatment and control groups received a battery of pretests and posttests. One of the tests was a valid and reliable author-constructed achievement test covering the minicourse content. The validity was judged by a panel of science educators and neuroscientists. Two forms of the test were developed initially and item analyses were used to determine items in the final, reliable form. The other two tests were the State section of the State-Trait Anxiety Inventory (STAI) - Form X (Spielberger, 1968) and the Test Anxiety Scale (Sarason, 1972). In addition, baseline data were gathered on student reaction to the course, e.g., what they liked, disliked, and believed. Teacher reaction was solicited by means of informal questionnaires and comments. It is important to emphasize that the teachers in the study were volunteers and their own classes were the student groups used. The volunteer element was deemed necessary in order to control for teacher stress in a new situation and, from a practical standpoint, to gain access to real-life classroom situations.

RESULTS

The primary statistical test used to analyze the data was Analysis of Covariance (ANCOVA) with pretest as the covariance and the posttest as the criterion. With respect to achievement test scores, group A was significantly higher than both groups B and C (the control group); group B was significantly higher than Group C. Although no significant differences were found among groups when comparing anxiety levels, the lack of linearity of the pre- and post-TAS regression lines led to use of a quadratic slope model and generated evidence of a trend in the expected direction at the higher anxiety levels. (See Figure 2).

Both student and teacher perceptions of the course were positive. The results of the student questionnaire provide data about reaction to the activities and course format. It was encouraging to note than 86% of the students felt that they knew more about bodily reactions during stress as a result of their exposure to the course and text anxiety coping treatment; 89% of the students thought that people ought to know what to do to help *themselves* in time of stress. A surprising 40% of the students said they would read more about stress even though the course was over!

In the area of course format, all six teachers participating in the study found the words and pictures in the minicourse booklet to be appropriate for their students. Additional comments included reference to a course on family life stress which the students of one teacher requested and which was handled successfully by a team of psychologists; the course was a direct result of student reaction to the stress minicourse. The teachers were satisfied with the information provided in the teacher's guide.

FOLLOW-UP STUDIES

Further work with the stress minicourse has ensued in a northern Florida school system and in a south Georgia private school. The Florida endeavor was sponsored by the local Mental Health Association and took place during

Figure 1

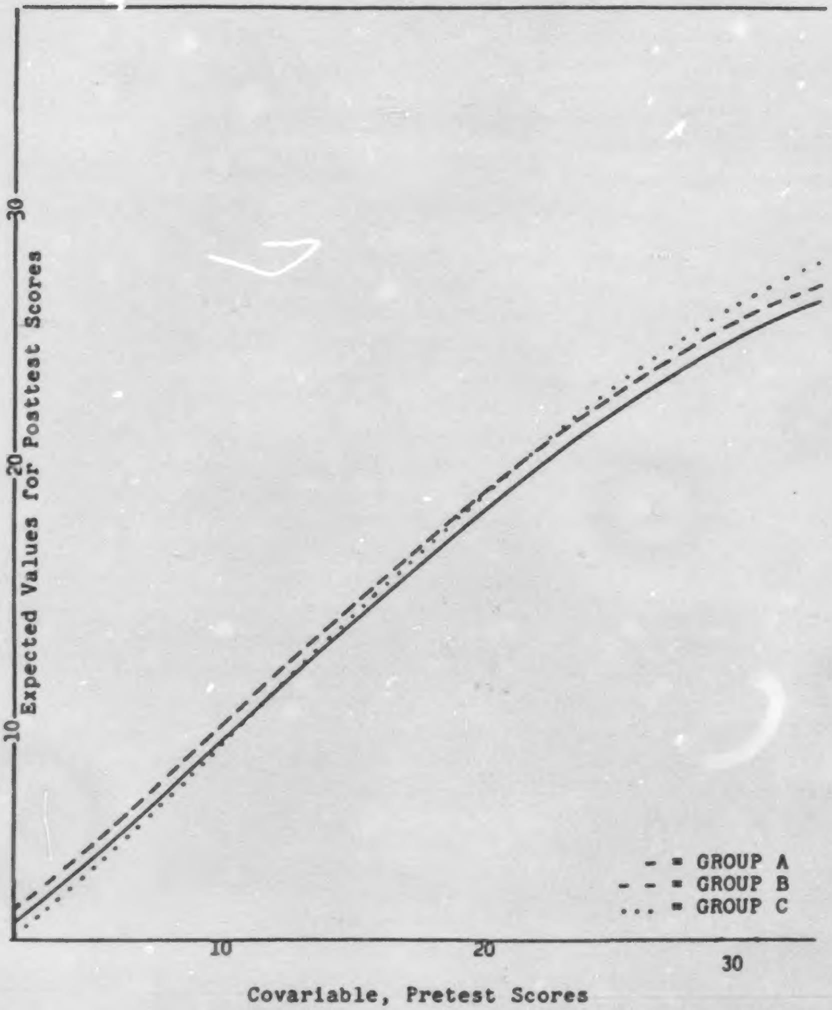
SCIENCE FRAMEWORK MATRIX

WHAT PEOPLE DO

	-A- Living with Self	-B- Career Choices	-C- Consumer Choicese	-D- Leisure Choices	-E- Living with Others (Culture)
1 Values, Attitudes, Aspiration	A-1 Belief in Myself	B-1 Commitment	C-1 Consideration of Consequences	D-1 Worthy Choice	E-1 Concern for People
2 Process Skills (Rational)	A-2 Organize Thought Processes	B-2 Decision Making	C-2 Product and Service Selection	D-2 Logical Decisions	E-2 Constructive Adjustment
3 Psychomotor (Manipulative Skills)	A-3 Competence in Handling Materials	B-3 Use of Basic Materials	C-3 Mechanical Aptness	D-3 Creative Application	E-3 Safe Practices
4 Knowledge	A-4 Mental and Physical Well-Being	B-4 Choice and Competence	C-4 Technical Application	D-4 Hobbies and Recreation	E-4 Interaction of Science, Technology & Society

Figure 2

TAS PRE- POST-REGRESSION LINES



two different academic years. In addition to the course, a group of mental health professionals interacted with the students in small group settings to provide information and answer student questions. Data collected in Florida helped the author of one of the anxiety scales to generate norms for high school students. Student questionnaire results were simular to those in the Texas study. The Mental Health Association viewed the venture as a successful growth experience for students, teachers and mental health professionals! Students perceived the concern of the outside speakers and reacted positively to an often neglected area of education — community involvement.

The Georgia experience involved a direct course offering by the author in cooperation with the science teacher during a special "mini-semester" time of the year in which students could register for courses of choice. Approximately 20 students elected the course. Students felt free to express opinions and to speak about their own personal sources of stress. An encouraging result was the students' openness and their asking the "stress professor" to return. Two other teachers in the school also expressed interest in a stress seminar and said they felt that they spoke for others, also. There was evidence that science interest in general was generated by the course, judging from student questions during the course and the teacher's follow-up comments.

CONCLUSIONS

With respect to achievement, the results indicate that tenth-grade biology students were able to learn the scientific aspects of emotional stress. While these cognitive gains are critical, affective gains must also be considered in a course based on the rationale established earlier in this paper. In effect, students enjoyed the course and teachers enjoyed teaching it. Students expressed belief in the premise that they themselves must take responsibility for solving their own problems. How universal that belief is needs to be researched. However, in a time when living sanely by understanding, predicting, and influencing our milieu is shaky at best, that conclusion is cause for plaudits.

A note of caution must be interjected here. In their zeal, curriculum planners are apt to assume that knowledge of a subject automatically leads to change in behavior. In effect, this is often not the case. This research suggests that knowledge of emotional stress alone does not automatically result in improved ability to cope with an actual situation involving anxiety prior to test-taking. However, there is evidence suggesting change in attitude, a necessary first step in problem-solving. The desire for instant results must give way to thoughtful planning and research design.

IMPLICATIONS

The program described in the paper has met with positive reaction among students, teachers, and science educators. The latter group was enthusiastic after discussions resulting from a presentation at the 1981 National Science Teachers Association annual meeting. It is the learning process which employs the scientific method — observation, experience, and revision that makes coping with stress a possibility and a probability. These very processes form the foundation of science, and when applied successfully to human behavior, they verify what Jacob Bronowski has called the "tolerance and humanity which must be present in the pursuit of science." (Bronowski, 1973). As science educators, we must express our own "humanness" and strengthen the link between science and survival. Our personal and professional lives, and those of our students, depend on it!

REFERENCES

Battle, J. Relationship between self-esteem and depression among high school students. *Perceptual and Motor Skills*, 1980, *51*, 157-158.

Bronowski, J. *The ascent of man.* Boston: Little, Brown and Company, 1973.

Brown, B. B. *New mind, new body — biofeedback: New directions for the mind.* New York: Harper and Row, Publishers, 1974.

Coping with depression. *Newsweek*, January 8, 1973, *81*, 51-54.

Gallup, G. H. The 13th annual Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan.* 1981, *63*, 34.

Johnson, S. O. Better discipline for middle school students. *The Clearing House*, 1979, *53* (2), 86-89.

Klopfer, L. E. Evaluation of learning in science, in Bloom, Hastings, J. & Madaus, F. *Handbook on Formative and Summative Evaluation of Student Learning*, New York: McGraw-Hill Book Co., 1971.

Richardson, F. C. *Coping with test anxiety: A guide.* The University of Texas at Austin, 1973.

Sarason, I. G. Experimental approaches to test anxiety: Attention and the uses of information. In Spielberger, D., ed. *Anxiety: current trends in theory and research, vol. II*, New York: Academic Press, 1972.

Saylor, G. J., Alexander, M. A. & Lewis, A. J. *Curriculum planning for better teaching and learning* (4th ed.), New York: Holt, Rinehart and Winston, 1981.

Shane, H. G. A curriculum for the new century, *Phi Delta Kappan.* 1981, *62*, 351-356.

Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. *The state-trait anxiety inventory (STAI), test manual for form X*, Palo Alto, California: Consulting Psychologists Press, 1968.

Texas Education Agency Science Section. *Science framework K-12, Document No. CU0 84101.* Austin, Texas: Texas Education Agency, September, 1974.

EFFECTS OF SCIENCE INQUIRY ACTIVITIES ON THE DEVELOPMENT OF CLASSIFICATION AND COMMUNICATION SKILLS OF CHILDREN LEARNING A SECOND LANGUAGE

IMELDA Z. RODRIGUEZ
BILINGUAL EDUCATION OFFICE
AUSTIN INDEPENDENT SCHOOL DISTRICT
AUSTIN, TEXAS

INTRODUCTION

One of the major goals of science education is the development of formal or abstract thought. Formal or abstract thought is defined as the ability to make observations and inferences, form hypotheses, classify and predict. Most of the thinking skills, according to the literature, are defined as inquiry or intellectual skills (Bruner, 1960; Suchman, 1961; Gagne, 1970). These inquiry skills are necessary tools for science learning. Classifying is of particular importance because it identifies and isolates variables, and is a precursor to reading development. Science educators supporting this view are Renner and Ragan, 1968; George and Dietz, 1969; Bethel, 1974; George, Dietz, Abraham & Nelson, 1974; Nelson, 1978.

Science through an inquiry approach presents many opportunities for children to discuss and communicate with one another. This helps achieve a major goal of language teaching — the development of oral communication skills (Hymes, 1967; Oller & Obrecht, 1968; [illegible]astin, 1971; Carroll, 1974).

The challenge for educators, then, is to create environments in the school that are conducive to languge learning and the development of intellectual skills. This kind of setting is important for the second language learner (L2) first, because the student is provided an opportunity to hear and practice language in context, and second, because the verbal interactions serve a communicative function (Fishman, 1974).

PURPOSE

The purpose of this investigation was to determine the effectiveness of an inquiry approach to science and language teaching to further develop classification and oral communication skills of Mexican-American third-grade children enrolled in a bilingual education program. Inquiry approach as used in this study has the following characteristics:

1. Manipulative materials are available to every student during each lesson.
2. The teacher interacts with the children.
3. The children work independently and also in small groups.
4. Inquiry mode of learning whereby children discover relationships through their own activity is emphasized.

RESEARCH HYPOTHESES

The two major research hypotheses were:

1. Will there be a difference in the mean total performance scores in classification skills between those subjects participating in the science inquiry lessons and those not participating?
2. Will there be a difference in the mean total performance scores in oral communication skills between those subjects participating in the science inquiry lessons and those not participating?

RESEARCH QUESTIONS

1. Does participation in thirty science lessons enhance classification skills?
2. If classification skills are enhanced through the inquiry lessons, which of the modes of classifying are enhanced: concrete, functional, or conceptual classification?
3. If classification skills are enhanced, is there a significant difference in the performance between male and female subjects?
4. If classification skills are enhanced, is there a significant difference in performance between those subjects that took the pretest and those that did not take the pretest?
5. Does participation in thirty science lessons enhance oral communication skills?
6. If oral communication skills are enhanced, which of the following subskills are enhanced: language output and expressiveness, vocabulary, general meaning and ideas, sentence structure, defining words, average length of sentence, listening behavior or listening comprehension?
7. If oral communication skills are enhanced, is there a significant difference in performance between those subjects that took the pretest and those that did not take the pretest?

RESEARCH DESIGN

POPULATION

This study was conducted in an urban elementary school, located in a central Texas school district. The student population consisted of 64% Mexican-American children, 35% black, and 1% Anglo-American. This school is a Title I school, which means the majority of the students are from low-income families.

The population consisted of third-grade Mexican-American bilingual subjects whose test scores in the English version of the Primary Acquisition of Languages (PAL) test were below 85 and whose scores on the Spanish PAL were above the English score. These children had been participating in a federally funded Title VII bilingual education program science kindergarten, but as they learned English they progressed toward an English language curriculum. They understand, speak, read, and write both Spanish and English, but are not sufficiently fluent in English to participate in a basic English language curriculum. This was the reason for participation in a bilingual education program. The opportunity for these children to interact with English-speaking peers was limited to the school; therefore, it was important to provide an environment that facilitated oral communication in the classroom.

SAMPLE

Sixty-four subjects were randomly selected from a population of 120 third-grade Mexican-American students. The composition of the third-grade consisted of 1% Anglo-American, 15% black, and 84% Mexican-American. The sample consisted of 16 males and 16 females in both the experimental and control groups. The age of the subjects ranged from 8.3 to 9.4 years and the mean age was 8.7.

The Solomon Four-Group Experimental Design (Campbell & Stanley, 1963) was used to assess the effects of 30 science inquiry lessons on the development of classification and oral communication skills as measured by two tests — the *Goldstein-Sheerer Object Sorting Test (GSOST)* and the *Test of Oral Communication Skills (TOCS).*

TREATMENT

The experimental group participated in a series of 30 sequential inquiry lessons (Boyd, 1970; Boyd & George, 1973; Bethel, 1974). The inquiry science lessons were selected because they are designed for the purpose of developing classification and oral communication skills. They are arranged in order of hierarchy so that concepts are introduced in order of increasing complexity. The lessons require the use of concrete objects, e.g., rocks, buttons, blocks, leaves, and shells. The children manipulate these objects and are required to make observations, comparisons, descriptions, and to form classes. The activities allow children to discuss their experiences with their peers and with their teachers.

It was through these discussions that the children were to develop oral communication skills. Each lesson had specific terms introduced such as "solubility," "texture," and "permeability." Therefore, children learned vocabulary in context, meanings of words, and operational definitions of terms. In addition, children learned to express themselves in precise and descriptive language since this was the nature of the science language used.

The students assigned to the experimental group were placed in two separate classrooms and were assigned to one teacher for their science instruction. The study was conducted for a period of 12 weeks, and it included the time required for testing. Each lesson was approximately 45 minutes long and was conducted daily for a period of 12 weeks.

Table 1
SOLOMON FOUR-GROUP DESIGN

I.	Experimental:	0_1 x 0_2	
II.	Control:	0_3 0_4	
III.	Experimental:	x 0_5	
IV.	Control:	0_6	

Total Sample Size

N=64 subjects

cell size

I. = 16
II. = 16
III. = 16
IV. = 16

The symbol 0 represents the administration pre- and post-assessment instruments: Goldstein-Sheerer Object Sorting Test (GSOST) and Test of Oral Communication Skills (TOCS). The symbol x represents the administration of the experimental treatment: Science Inquiry Lessons.

The control group was assigned to two different classrooms and participated in a science program that did not emphasize classification and oral communication skills development specifically nor require the manipulation of concrete objects. The science program used with the control group consisted of units developed by teachers working in the identified school district. This program consisted of units of study on different topics, e.g., "the Frog," "the Fish," and "Energy."

Prior to the first lesson, one-half of the subjects in each control and experimental groups were pretested using the two instruments: *Test of Oral Communication Skills (TOCS)*, which measures the child's ability to communicate orally, and the *Goldstein-Sheerer Object Sorting Test (GSOST)*, which measures the child's ability to classify concrete objects. At the completion of the study all 64 students were posttested using the same instruments.

INSTRUMENTATION

The *Goldstein-Sheerer Object Sorting Test (GSOST)* as adapted by Rapaport, Gill, Schafer (1968) uses familiar objects, identifies the modes of classifying (concrete, functional, and operational), measures active sorting (initiated by the classifier), and uses passive sorting (initiated by the investigator). The test is administered on an individual basis and takes approximately 20 minutes to administer.

The *Test of Oral Communication Skills (TOCS)* developed by Phyllis Huff (1973) is designed specifically to measure oral language skills through experiences acquired in a science program. These skills include language output and expressiveness, vocabulary, general meaning and ideas, sentence structure, defining words, average length of sentences, listening behavior, and listening comprehension.

RESULTS

Analysis of variance ANOVA was the statistical technique used for all analyses (Helwig, 1978). In addition to the main treatment effects, the following factors were considered: (1) pretest learning effects and (2) differences in performance between male and female subjects.

A significance level of 0.01 was used for all analyses. Analysis of the posttest mean scores revealed that participation in the science inquiry lessons did result in improved classification and oral communication skills for the experimental groups. No significant results were obtained from analysis of pretest learning effects and no differences were found in performances between male and female subjects.

Analysis of the posttest mean scores revealed that participation in the science inquiry lessons did result in improved classification skills for the experimental groups (see Table 2).

Table 2

COMPARISON OF *CLASSIFICATION* POSTTEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS

	df	ss	F	p
Between Groups	1	293.26	13.4	0.005

	N	M.S	S.D
Experimental Group	32	20.06	4.08
Control Group	32	15.75	5.25

The analysis showed a significant difference in posttest scores on the GSOST, but since this test measures three different modes of classifying, analysis of each mode was performed sequentially. Since the test is divided into two parts and each part has a different format, an analysis of each part of the test was performed.

The analysis further indicated that significant differences occurred in the conceptual mode of classifying in both parts of the test. Concrete and functional modes were not affected significantly in either part of the test. These results indicated that the inquiry science lessons significantly improved the conceptual mode of classifying which is the highest level of classification (see Tables 3 and 4).

Table 3

COMPARISON OF *PART I CLASSIFICATION* POSTTEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON EACH MODE OF CLASSIFYING

Source Part I	df	ss	F	p
Concrete	1	9.00	8.89	0.041
Functional	1	0.765	0.76	0.3853
Conceptual	1	8.26	27.75	0.0001

Mode	N	M.S	S.D
Concepteual (E)	32	.72	.77
Conceptual (C)	32	0.25	0.56

Table 4

COMPARISON OF *PART II CLASSIFICATION* POSTTEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON EACH MODE OF CLASSIFYING

Source Part II	df	ss	F	p
Concrete	1	6.8	.89	0.3502
Functional	1	5.64	0.81	0.3720
Conceptual	1	23.76	14.79	0.0003

Mode	N	M.S	S.D
Concepteual (E)	32	1.9	1.48
Conceptual (C)	32	0.68	0.99

Examination of oral communication posttest scores revealed a significant difference between the experimental and control groups (see Table 5).

Table 5

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS

	df	ss	F	p
Between Groups	1	1806.25	71.98	0.0001

	N	M.S	S.D
Experimental Group	32	40.31	5.18
Control Group	32	29.68	4.83

The TOCS is divided into two parts: Part One measures oral transmission skills such as language output and expressiveness, vocabulary, general meaning and ideas, sentence structure, defining words, and average length of sentence; Part Two measures oral receiving skills which include listening behavior and listening comprehension. An analysis of each subskill was performed to determine which subskill or subskills were affected.

The results indicated that significant differences occurred in each subskill; therefore, it can be concluded that total oral communication skills are enhanced by participation in the science inquiry lessons (see Table 6 and 7).

Table 6

TOCS MEAN SCORES AND STANDARD DEVIATIONS FOR EACH SUBSKILL

Skill	N	M.S	S.D
Language Output	32 (E)*	4.09	.69
	32 (C)	2.93	.61
Vocabulary	32 (E)	3.75	.625
	32 (C)	2.56	.59
General Meaning	32 (E)	3.985	.72
	32 (C)	2.84	.590
Sentence Structure	32 (E)	3.967	.699
	32 (C)	2.93	.566
Defining Words	32 (E)	3.9	.459
	32 (C)	2.74	.54
Sentence Length	32 (E)	3.967	.643
	32 (C)	2.93	.41
Listening Behavior	32 (E)	7.90	1.318
	32 (C)	6.00	1.435
Listening Comprehension	32 (E)	8.77	1.33
	32 (C)	6.58	1.13

* (E) Experimental; (C) Control

Table 7

COMPARISON OF TOCS POSTTEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON *EACH SUBSKILL*

Source	df	ss	F	p
Language Output	1	21.39	49.87	0.0001
Vocabulary	1	22.56	58.59	0.0001
General Meaning	1	19.14	42.24	0.0001
Sentence Structure	1	16.0	50.23	0.0001
Defining Words	1	23.7	59.01	0.0001
Sentence Length	1	16.0	72.15	0.0001
Listening Behavior	1	58.14	29.37	0.0001
Listening Comprehension	1	81.0	48.55	0.0001

ANOVA of posttest classification scores of both control and experimental groups revealed no significant differences between those pretested and those not pretested.

ANOVA of posttest oral communication scores revealed no significant differences between the control and experimental groups pretested and not pretested.

ANOVA of posttest scores revealed no significant differences between males and females in both oral communication and classification skills.

The nonsignificant results of the pretest analysis confirmed a basic assumption of the study that both groups were equivalent in classification skills and oral communication skills. Therefore, it can be concluded that both groups were randomly drawn from the same population.

PRETEST RESULTS

ANOVA of pretest mean scores on both parts of the GSOST revealed no significant differences between the control and experimental groups. Similar results were found between the control and experimental groups on the TOCS test of oral communication skills.

DISCUSSION

Classification skills of bilingual children, particularly the conceptual (abstract) mode which is the highest level of classifying, did improve as a result of the experience. These conclusions are compatible with studies on deaf children

(Boyd, 1970; Boyd & George, 1973) and black inner city children (Bethel, 1974).

An improvement in oral communication skills was found. This was expected because a major emphasis of this investigation was oral communication. Children were given concrete objects and tasks which involved manipulation of these objects, exploration, and verbal interaction. These experiences provided the children with opportunity to talk about something they were constantly doing. They discussed these experiences among themselves and with their teacher on a regular basis. The quality of oral language increased because they had something different or new to talk about over an extended period of time.

It can be concluded that participation in inquiry science lessons had an overall positive effect on the oral communication and classification skills of Mexican-American bilingual children. It may be that the linguistic objectives of these lessons were in some way related to the cognitive skills of science and were therefore appropriate for these children as well as most other children who were evaluated and tested in related studies.

Teaching science as inquiry with an emphasis on concrete experiences appears to be a viable means of developing both classification and oral communication skills. Methods such as those used in this study are recommended for use with children in need of oral communication and classification skills development. The results obtained here were accomplished in 12 weeks. The possibilities of teaching science in this manner over an extended period, e.g. nine months, to children could have profound implications for both their intellectual development as well as their reading skills development. It is hoped that methods such as these will be used by other teachers with other children.

IMPLICATIONS

A unique aspect of science inquiry lessons used in this study to teach science and language is that both tasks are accomplished simultaneously. This study addressed Mexican-American children who are in the process of learning English as a second language. The positive significant results found in this study may have implications for teaching science in Spanish to children who are in the process of learning Spanish as a second language.

Language learning, whether it be English, Spanish, Vietnamese, or any other language, can be developed through teaching science as inquiry with concrete objects. This kind of methodology provides the means for communication in the new language because it is interesting and motivational. It also provides the framework necessary to develop the grammatical structures of the new language. Supporting this view of second language pedagogy are Rivers, 1975; Paulston, 1974; Rubin, 1975; Wong-Filmore, 1976; Matluck, 1977; Sampson, 1977; and Edelsky, 1978. It is therefore strongly recommended that science inquiry be included in the curriculum of bilingual education and English as a Second Language program.

REFERENCES

Barufaldi, J. P., & Swift, J. W. The influence of the BSCS-elementary school sciences program instruction on first-grade students' listening skills. *Journal of Research in Science Teaching, 17* (5), 1980, 485-490.

Bethel, L. J. Science inquiry and the development of classification and oral communication skills in innercity children. (Doctoral dissertation, University of Pennsylvania, 1974). *Dissertation Abstracts*, 1974, 75-037-02.

Boyd, M. E. A study to determine the effects of science inquiry on the abstract categorization behavior of deaf children. (Doctoral dissertation, University of Pennsylvania, 1970). *Dissertation Abstracts*, 1974, 74-137-65.

Boyd, E., & George, K. D. The effect of science inquiry on the abstract categorization behavior of deaf children. *Journal of Research in Science Teaching, 10* (1), 1973, 91-99.

Bruner, J. S. *The process of education.* New York: Random House, 1960.

Campbell, D. & Stanley, J. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1963.

Carrol, J. Learning theory for the classroom teacher. In G. A. Jarvis (Ed.), *The challenge of communication.* Skokie, Illinois: National Textbook Company, 1974.

Chastin, K. The development of modern language skills. *Theory to practice.* Philadelphia: Center for Curriculum Development, 1971.

Edelsky, C. Teaching oral language. *Language Arts, 55* (3), 1978.

Fishman, J. A. A sociology of bilingual education. *Final Report to the Division of Foreign Studies.* Washington, D.C.: Department of Health, Education and Welfare, Office of Education, September 1974, (OECO-73-0588).

Gagne, M. & Brown, L. Some factors in the programming of conceptual learning. *Journal of Experimental Psychology, 62*, 1961, 313-321.

George, K. D., & Dietz, M. A. An experimental comparison of urban children with suburban children on eight tasks of basic skills. *Journal of Research in Science Teaching, 6*, 1969, 224-233.

George, K. D., Dietz, M. A., Abraham, E. C., & Nelson, M. A. *Elementary school science: Why and how.* Lexington, Massachusetts: D. C. Heath, 1974.

Helwig, J. T. *Statistical analysis system (SAS).* Raleigh, North Carolina: SAS Institutes, Inc., 1978.

Huff, P., & Lanugis, M. The effects of the use of activities of SAPA on the oral communication skills of disadvantaged kindergarten children. *Journal of Research in Science Teaching, 10* (2), 165-173.

Hymes, D. The anthropology of communications. In F. Dance (Ed.), *Human Communication Theory.* New York: Holt, Rinehart, and Winston, 1967.

Matluck, B. The order of acquisition of certain oral English structures by native speaking children of Spanish, Cantonese, Tagalog, and Ilokano learning English as a second language. Between the ages of five and ten. (Doctoral dissertation, The University of Texas at Austin, 1979.) *Dissertation Abstracts*, 1977, 77-229.91.

Nelson, K. *Children's language.* New York: Garden House, *1*, 1978.

Oller, J., & Obrecht, D. J. Pattern drill and communicative activity: A pscholinguistic experiment. *International Review of Applied Linguistics, 6*, 1968, 165-174.

Paulston, C. B. Linguistic communicative competence. *TESOL Teachers of English to Speakers of Other Languages Quarterly, 8* (4), 1974

Piaget, J., & Inhelder, B. *The psychology of the child.* New York: Basic Books, 1969.

Renner, J., & Ragan, W. *Teaching science in the elementary school.* New York: Harper and Row, 1968.

Rivers, W. M. Talking off the tops of their heads. *Teachers of English to Speakers of Other Languages Quarterly, 9* (1), 1972.

Rubin, J. What the good language learner can teach us. *TESOL Quarterly,* 9 (1), 1975.

Sampson, G. P. A real challenge to ESL methodology. *Teachers of English to Speakers of Other Languages Quarterly, 3* (1), 1977.

Suchman, J. R. Inquiry training: Building skills for autonomous discovery. *Merrill-Palmer Quarterly,* 1961, 147-169.

Swift, J. The effects of participation in selected activities from the biological sciences curriculum study — Elementary school science program on the enhancement of listening skills. (Doctoral dissertation, The University of Texas at Austin, 1977). *Dissertation Abstract,* 1977, 77-230.36.

Vygotsky, L. S. *Thought and language,* Tr. Cambridge, MA: MIT press, 1962.

Wong-Filmore, L. The second time around: Cognition and social strategies in second language acquisition. (Doctoral dissertation, Stanford, 1976). *Dissertation Abstracts,* 1976, 77-045-62.

OBJECTIVES AND PLANNING CONCERNS FOR SECONDARY SCIENCE METHODS COURSES: A DELPHI STUDY

MARCIA VIRGINIA dos SANTOS SILVA
UNIVERSIDADE FEDERAL de SANTA MARIA
SANTA MARIA, BRASIL

INTRODUCTION

The study was designed to identify major objectives of science methods courses offered to prospective secondary science teachers and to determine what philosophical bases (or concerns), if any, underlie the choice of objectives for methods courses. Specifically the study sought to do the following: (1) seek and compare the opinions of university professors of science education about priorities for planning outcomes for students enrolled in secondary science methods courses, (2) determine the degree of importance and consensus existing among those experts regarding the different objectives and concerns which they have identified, and (3) determine whether instructors with different planning concerns consider certain objectives to be more important than others. Previous investigations have solicited from panels of experts opinions concerning the skills needed by practicing science teachers. Until now no one has polled methods instructors to see which objectives they think should be stressed in a methods course and why.

BACKGROUND

A science methods course should be one of the most significant courses in the preparation of a prospective science teacher. Indeed, it is supposed to provide prospective teachers with the strategies and tactics for teaching science, knowledge of pedagogical concepts and principles, and should develop in teachers certain attitudes toward science and science teaching. There is evidence to indicate that the methods course is not fulfilling its purpose. Butts (1968, p. 2) summarized all too well what appears to be a prevalent attitude, "Discussion of a science methods course is many times similar to that of the weather. There is a predilection for much discussion and little action."

Although the importance of a methods course is recognized by all who in any way are connected with teacher preparation, the number of published reports in which specific recommendations are made is limited. The studies available tend to be isolated efforts reflective of the investigator's philosophy, interests, and beliefs on values concerning science teaching. For example, Anderson (1969) maintains that a science methods course should help prepare the prospective teacher to assume a role in an "updated" teaching situation, aware of the immediate and projected needs of science education and knowledgeable of the mechanism of change and reform. Druger (1969) stresses teaching self-awareness to prospective teachers. Bicak (1975) recommends the laboratory approach be used by science teachers in secondary schools and offers recommendations concerning the nature of experiences needed in the methods course. Golman (1975) emphasizes 75 tasks for students enrolled in the methods course. Each educator reports much success in helping students attain course objectives, objectives reflecting much understanding and insight into the numerous tasks science teachers perform in their schools.

Several investigations sought help from panels of experts to identify the skills needed by secondary science teachers in order to be effective teachers, presumably for the purpose of guiding professional program development. Using the California definition of a competent teacher, Spore (1962) sought to identify competencies needed by secondary science teachers to fulfill their roles in promoting student growth, acting as a liaison to the community, and building an educational program. Professors of science education, science supervisors, and science teachers rated 60 predetermined competencies. Science teachers and supervisors were in agreement on the needed competencies, but their lists disagreed considerably with that identified by science educators. More than a decade later Quershi (1975) also found disagreement between science teachers and science teacher educators concerning essential competencies. Of the 12 competencies produced by science teachers, science teacher educators agreed on 5 (only 2 competencies are included in Spore's list). In 1977 Simpson and Brown called upon educators in Georgia and Colorado to validate a set of 23 basic competencies for teachers of secondary school science, competencies produced in 1974 by a committee composed of representatives of the Association for the Education of Teachers in Science and the National Science Teachers Association. As a result of their study, Simpson and Brown identified seven fundamental skill areas for science teachers: knowledge of science, professional knowledge and attitude, human relations skills, planning skills, instructional skills, management skills, and evaluation skills. Using the Delphi technique Chiappetta and Collette (1978a) obtained consensus among science supervisors regarding eight skills needed by secondary science teachers. These eight skills correspond closely to the seven fundamental areas identified in the Simpson and Brown study.

The concerns of preservice teachers offer valuable insights into the frame of reference of the prospective secondary science teacher, information which can be extremely helpful to instructors planning science methods courses. Fuller (1979) posed a three phase developmental conceptualization of teachers' concerns after examining transcripts of discussions held with preservice teachers and written responses about their concerns at the time. Preservice teachers, according to Fuller, are first concerned about themselves (Self Concerns), later about teaching tasks (Task Concerns), and finally about the impact of their teaching upon students (Impact Concerns). Spivey (1976) offers a similar model to explain the development of successful teachers upon entering the profession. Understanding that preservice teachers progress through predictable levels of concern in the process of becoming a professional secondary science teacher offers methods instructors one basis for identifying objectives and sequencing learning experiences for the science methods course.

71

RESEARCH DESIGN

The study was designed to provide answers to three research questions.

(1) Is there agreement among university science educators about what a preservice secondary science teacher should be able to do upon completing a methods course?

(2) What are the philosophical bases (or concerns) underlying the choice of objectives for a secondary science methods course?

(3) Do instructors with different concerns consider some objectives to be more important than others?

The Delphi Technique has been successfully used with knowledgeable persons in many fields, including education, to reach consensus on matters about which they were well informed. This technique was used in this study to gain consensus about the priorities of university science educators for planning a secondary science methods course. A complete description of the Delphi technique is given by Helmer (1967).

PROCEDURE

Thirty-three science educators were solicited to be members of the panel of experts for the study. They were selected from a list of faculty members of institutions throughout the United States with graduate programs in science education. Although these institutions are not representative of all institutions preparing secondary science teachers, their preparation programs are among the most visible and systematically studied. Of the 33 educators solicited, 27 (81.8%) agreed to participate in the study and 24 (88.8%) of them completed it. The 24 participants represented 23 institutions with graduate programs in science education located in 16 states throughout the U.S. The sample size of 24 was comparable to that used by Simpson and Brown (1977) and considerably larger than the five experts used by Stolovitch (1977) to forecast trends in teacher education. The procedure used to gather information from the panel of experts followed that suggestion by Pfeiffer (1968, pp. 152-153) except that a fourth round of questionnaires was excluded, since previous researchers (Butts, Capie, Fuller, May, Okey, & Yeany, 1977; Clark & Coutts, 1971; Cyphart & Gant, 1971), have shown that convergence is greatest on the first round following feedback.

Three rounds of questionnaires were used in the study. Round one consisted of two questionnaires, one an informational questionnaire (in two parts), the other demographic. The informational questionnaire consisted of forced and free response items in Part I which was designed to identify objectives that each panel member believed to be important outcomes for students enrolled in a secondary science methods course. Panel members were presented with 10 simple behavioral objectives (lacking the condition and criterion of performance) and were asked to add any additional objectives that they thought represented worthwhile things a student should be able to do upon completing a methods course. In Part II panel members were asked to describe the philosophical basis underlying their choice of objectives for students enrolled in a secondary science methods course, the focus of their concern when planning the course.

The demographic questionnaire provided information about panel members' professional experience and institutional responsibilities. Respondents estimated the percent of time they teach, perform administrative duties, and conduct research as part of their contractual obligation to their respective institutions. In addition, information was obtained about each person's experience teaching science in a secondary school, teaching secondary science methods courses, and working with preservice secondary science teachers.

Information obtained from an analysis of the responses to Parts I and II formed the basis for writing objectives and planning concerns used in the first Delphi questionnaire (round 2) and the second consensus-seeking Delphi questionnaire (round 3). Data from the final questionnaire were collected and analyzed for the degree of consensus existing among panel members. The collection of data covered approximately four months, May through August, 1979.

RESULTS

Panel members were found to vary considerably in their university responsibilities and their experience as science educators. Of the 24 educators only two had single university responsibilities: one taught full time; the other's time was devoted to administrative responsibilities. At the other extreme, 14 educators indicated that they had three university responsibilities: eleven teaching, administration, and research; three teaching, research, and "other." Of the 24 panel members, 23 indicated they had taught secondary science. All panel members had experience as university science educators (from 7 to 35 years) and had been teaching or administering methods courses for prospective secondary science teachers (from 1 to 35 years). Among the institutions from which the panel members were drawn there seems to be fairly widespread agreement about when students should enroll in the methods course and student teaching: 17 members reported that the methods course precedes student teaching; two reported that it could be taken before or during but not after student teaching; and four reported that the methods course could be taken anytime prior to completing the professional program.

Fifteen objectives and five general planning concerns were identified from the Round One Informational Questionnaire. The objectives formed Part I and the planning concerns formed Part II of subsequent (rounds 2 and 3) Delphi questionnaires.

As a result of the Round Two Delphi Questionnaire, consensus was reached on each of the 15 objectives, though at differing levels of importance. On a scale of 1 (very unimportant) to 5 (very important) the objective "Plan one or more science lessons" was rated highest with an average rating of 4.91. Rated the lowest was "Design remedial activities for students who lack prerequisite skills for a given set of objectives" with a rating of 3.56. Table 1 contains data gathered from Part I of the Round Two Questionnaire: a listing

Table 1

ANALYSIS OF RESPONSES TO ROUND THREE DELPHI QUESTIONNAIRE: PART I (N = 24)

OBJECTIVES/RATING*	1		2		3		4		5			Standard	% Above Or
	No.	%	No.	%	No.	%	No.	%	No.	%	Mean	Deviation	Below 3
1. Teach one or more science lessons.	1	4.2	0	0.0	1	4.2	3	12.5	19	79.1	4.62	.92	+ 92
2. Make a collection of teacher resource materials for a science area.	1	4.2	1	4.2	1	4.2	16	66.7	5	20.8	3.95	.90	+ 88
3. Design a plan for incorporating safety procedures into all aspects of science classroom teaching	0	0.0	2	8.3	3	12.5	11	45.8	8	33.3	3.90	.88	+ 79
4. Plan an entire laboratory investigation.	0	0.0	1	4.2	1	4.2	5	20.8	17	70.8	4.58	.77	+ 92
5. Plan one or more science lessons.	0	0.0	0	0.0	0	0.0	2	8.3	22	91.7	4.91	.28	+100
6. Describe ways to provide for unusual emotional and physical needs of students when planning science lessons.	0	0.0	0	0.0	10	41.7	12	50.0	2	8.3	3.66	.63	+ 58
7. Design remedial activities for students who lack prerequisite skills for a given set of objectives.	1	4.2	2	8.3	5	20.8	13	54.2	3	12.5	3.56	.94	+ 67
8. Construct a long range teaching (unit) plan.	0	0.0	1	4.2	2	8.3	7	29.1	14	58.3	4.41	.82	+ 87
9. Construct a classroom evaluation scheme.	0	0.0	1	4.2	1	4.2	5	20.8	17	70.8	4.58	.77	+ 92
10. Describe interpersonal skills and strategies that are effective in dealing with adolescent behavior.	1	4.2	0	0.0	3	12.5	10	41.7	10	41.7	4.16	.96	+ 83
11. Describe the implications of science for science teaching.	0	0.0	0	0.0	3	12.5	10	41.7	11	45.8	4.33	.70	+ 88
12. Analyze the results of student attainment.	0	0.0	0	0.0	1	4.2	6	25.0	17	70.8	4.66	.56	+ 96
13. Design procedures for obtaining information as to the effectiveness of your teaching.	1	4.2	0	0.0	3	12.5	11	45.8	9	37.5	4.12	.94	+ 83
14. Explain the role of a science teacher as a professional.	0	0.0	2	8.3	2	8.3	16	66.7	4	16.7	3.91	.77	+ 83
15. Describe the role of learning theories in science teaching.	1	4.2	2	8.3	4	16.7	7	29.1	10	41.7	3.95	1.16	+ 71

* 1 = Very Unimportant; 5 = Very Important

of the 15 objectives, a summary of the responses, and the percent of the responses above or below three, which was needed to arrive at the degree of consensus. Panel members did not change their ratings of the 15 objectives of Round Three.

The degree of consensus was determined by computing the percent of ratings above three, if the average rating was above three, and the percent below three if the average rating was below three. Panel members expressed *very high* consensus (90% or more) on five objectives (1, 4, 5, 9, and 12 in Table 1), *high* consensus (80%-89%) on five objectives (2, 8, 10, 11, 13 and 14), *considerable* consensus (70%-79%) on two objectives (3 and 5), and *some* consensus (60%-69%) and *little* consensus (50%-59%) respectively on objectives 7 and 6.

In Part II of the Informational Questionnaire panel members were asked to respond in paragraph form to the question "When you think about planning a secondary science methods course, what are you concerned about?" Each statement in the paragraphs provided by panel members was written on an index card. The topic of each statement was identified and coded accordingly. Index cards were sorted by topic, and topics were grouped according to the similarity of their content. Using this technique of analyzing the content of each statement, five categories of planning concerns formed the basis of Part II for subsequent (rounds 2 and 3) Delphi questionaires.

From statements written by panel members the following five categories of planning concerns were identified:

1. *Teacher Concerns*
   This concern encompasses the interpersonal, methodological (planning, instructing, managing, evaluating, etc.), and attitudinal behaviors that a teacher must use in conjunction with an understanding of the learning process in order to teach science effectively and safely to high school students.
2. *Student Concerns*
   This concern supports the belief that instruction must be responsive to the differing characteristics, backgrounds, and needs (physical, emotional and achievement) that students bring to the science classroom.
3. *Subject Matter Concerns*
   This concern emphasizes the need for instruction to stress the philosophies, nature, processes concepts, and the diverse content of the subject area.
4. *Society Concerns*
   This concern reflects the need for the science teacher to serve as an interpreter of science for society and to help future citizens (students) acquire positive attitudes towards science and scientists.
5. *Curriculum Concerns*
   This concern supports the importance of teaching scientific knowledge as it relates to other fields of knowledge and of understanding the role science plays in the total high school curriculum.

Panel members ranked teacher concerns as the most important basis for planning a secondary science methods course. Student, subject-matter, and society ranked 2, 3, and 4 respectively as concerns that need to be stressed when planning the methods course. Science as it relates to other fields of knowledge (curriculum concerns) was ranked as the least important of the five planning concerns. A summary of the results to Part II of the final Delphi Questionnaire (Round Three) is found in Table 2.

Table 2

ANALYSIS OF RESPONSES TO ROUND THREE DELPHI QUESTIONNAIRE: PART II (N = 24)

	IMPORTANCE											
PLANNING CONCERNS	**1**		**2**		**3**		**4**		**5**			**Standard Deviation**
	No.	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**Mean**	
Curriculum	0	0.0	2	8.3	1	4.2	8	33.3	13	54.2	1.66	.91
Student	4	16.7	14	58.3	3	12.5	3	12.5	0	0.0	3.79	.88
Subject-matter	5	20.8	2	8.3	6	25.0	4	16.7	7	29.2	2.75	1.51
Society	1	4.2	2	8.3	8	33.3	9	37.5	4	16.7	2.45	1.02
Teacher	14	58.3	4	16.7	6	25.0	0	0.0	0	0.0	4.33	.86

* = Very Unimportant; 5 = Very Important

To determine whether instructors with different planning concerns consider certain objectives to be more important than others it was necessary to examine the final tally of responses of each panel member to Parts I and II of the Round 3 Questionnaire. Society and curriculum concerns, with the lowest average rankings were deleted from this analysis. No panel member ranked curriculum concerns as most important, and only one person attached that ranking to society concerns. Thus responses from 23 out of 24 panel members were used to investigate the relationship between rankings of the 15 objectives and instructors' major planning concerns.

Separating the 23 panel members into three groups according to their major planning concern revealed several differences in the average importance attached to the 15 objectives. Members of the "teacher" concerns group (N=14) rated all 15 objectives to be of greater importance, on the average, than did members of the "subject-matter" group (N=5). Comparing the ratings of "teacher" and "student" concerns groups revealed differences on 13 of the 15 objectives, the higher average ratings being recorded by members of the "teacher" group. The average rating from the "student" concerns group (N=4) exceeded that of the "teacher" concerns group on objective one ("teach one or more science lessons"); average ratings for the two groups were the same on objective 5 ("plan one or more science lessons"). No pattern was discerned when the average ratings of the "student" group were compared with those of the "subject-matter" group. The relative ratings of the 15 objectives within a group, as well as between groups, can be found in Table 3.

DISCUSSION

Science educators are in agreement, for the most part, with regard to what should be the outcomes of a secondary science methods course. Their primary

Table 3

SUMMARY OF RATING OF OBJECTIVES AND PLANNING CONCERNS FOR A SECONDARY SCIENCE METHODS COURSE

	PLANNING CONCERNS			
Methods Course Objectives	**Teacher (N = 14)**	**Subject Matter (N = 5)**	**Student (N = 4)**	**Average Rating of Objectives**
1	4.78	3.80	5.00	4.46
2	4.14	3.80	3.25	3.73
3	4.28	4.00	3.25	3.84
4	4.78	4.20	4.25	4.41
5	5.00	4.60	5.00	4.86
6	3.85	3.40	3.25	3.50
7	4.00	3.20	2.75	3.31
8	4.71	3.80	4.00	4.17
9	4.78	4.40	4.75	4.64
10	4.42	3.80	3.50	3.90
11	4.42	4.40	3.50	4.10
12	4.71	4.60	4.25	4.52
13	4.42	3.00	4.25	3.89
14	4.07	3.60	3.75	3.80
15	4.28	3.60	3.25	3.71
Average Rating by Concern	4.44	3.88	4.12	

NOTE: Society concerns omitted, N = 23 not 24 planning concerns differ; consequently, they emphasize objectives in varying degrees. Teacher concerns represent the primary focus of 14 of the 24 participants when they plan the secondary science methods course. It is reasonable to assume that this group played a major role in identifying the list of fifteen objectives on which consensus was sought. This is noteworthy. The "subject-matter" concerns group, people who believe the methods course should stress the philosophies, nature, processes, concepts, and the diverse content of science, were represented by so few people (N=5) that their contribution to the establishment of the original list of 15 objectives was minimal, perhaps, when compared with contributions made by the "teacher" concerns group (N=14). From the start the 15 objectives reflected "teacher" more than "subject-matter" outcomes. Lacking a representative list of objectives stressing "subject-matter" planning concerns may help to explain why this group registered a consistently lower average rating on each of the 15 objectives when compared with ratings of the "teacher" concerns group.

One outcome raises suspicion regarding the role that concerns play in the planning of a secondary science methods course. Four panel members identified the student to be their primary concern when planning a secondary science methods course for propective teachers. This concern reflects the belief that instruction must be responsive to the differing characteristics, backgrounds, and needs (physical, emotional and achievement) that students bring to the science classroom. Yet, the lowest rating given to any of the fifteen objectives by the three groups was given by the "student" concerns group to objective 7, "Design remedial activities for students who lack prerequisite skills for a given set of objectives." Such a finding raises some questions regarding the influence, if any, of planning concerns on methods course objectives.

Further study needs to be conducted. First, larger samples of science educators expressing primary planning concerns other than "the teacher" need to be identified. Holding major concerns for student, subject-matter, curriculum, or society could lead science educators to plan secondary science methods courses that focus on vastly different objectives. Each objective might be of major importance in preparing teachers to meet the variety of demands placed upon them in today's classroom. In addition, care must be taken not to destroy the wording of objectives offered by members of different concerns groups; in an effort to group objectives according to content or process, much valuable insight into the subtle differences in approach to setting course objectives may perhaps be lost.

As a result of this study future researchers need to adopt research strategies that yield more in-depth information about the philosophical and psychological concerns taken into consideration by university instructors when they plan secondary science methods courses. For example, interviews would permit the researchers to probe the underlying reasons for preferring one planning concern over another. Even without the use of in-depth interviews, additional information of value might be obtained through a carefully designed analysis of the content of the written comments provided on initial questionnaires, such as that collected on the informational questionnaire used at the beginning of this study.

Many educational groups have an interest in the nature of the training provided prospective secondary science teachers in methods courses. Do these groups share similar or quite different concerns about the role of the science methods course? This study demonstrates that even within a rather homogeneous group, a sample composed of university science educators, planning concerns differ and groups expressing different concerns vary in the average importance they attach to specific objectives for students enrolled in a secondary science methods course.

REFERENCES

Andersen, H. O. Science methods instruction. *The American Biology Techer, 31*(4), 1969, 250-251.

Bicak, L. J. Investigation as the content of a methods course. *Science Education, 59,* 1975, 345-349.

Butts, D. P. The science methods course of elementary education majors: a seminar. Seminar presented at the Southwest Meeting, Association for the Education of Teachers of Science, Austin, Texas, January, 1968.

Butts, D., Capie, W., Fuller, E., May, D., Okey, J., & Yeany, R. *Priorities for research in science education.* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Cincinnati, Ohio, March 23, 1977. (ERIC Document Reproduction Service No. ED 139 633).

Chiappetta, E. L., & Colletta, A. T. Secondary science teacher skills identified by science supervisors. *Science Education, 62*(1), 1978, 67-71. (a)

Chiappetta, E. L., & Colletta, A. T. Secondary science teacher skills identified by secondary science teachers. *Science Education, 62*(1), 1978, 73-78. (b)

Clark, S. C. T., & Coutts, H. T. *The future of teacher education.* Edmonton, Alberta, Canada: University of Alberta, 1971. (ERIC Document Reproduction Service No. ED 154 065).

Cyphert, F. R., & Gant, W. L. The delphi technique: A tool for collecting opinions in teacher education. *Journal of Teacher Education, 21*(3), 1979, 417-425.

Druger, M. An approach to a science methods course. *Science Education, 53*(5), 1969, 381-383.

Fuller, F. F. Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6,* 1969, 206-226.

Golman, M. E. Objectives for the science teaching methods course. *The Science Teacher, 42*(4), 1975, 21-22.

Helmer, O. *Analysis of the future: The delphi method* (RAND P-3558). Santa Monica, California: RAND Corporation, March, 1967.

Pfeiffer, J. *New look at education: Systems analysis in our schools and colleges.* New York: Odyssey Press, 1968.

Qureshi, Z. A practical approach to the identification of teacher competencies for high school science teachers (Unpublished doctoral dissertation, University of Maine, 1975), *Dissertation Abstracts International*, 1965, *37,* 903A-904A. (University Microfilms No. 76-17, 116).

Simpson, R. D., & Brown, D. R. Validating science teaching competencies using the delphi method. *Science Education, 61* (2), 1977, 209-219.

Spivey, J. R. A conceptual model of teacher development. *Kappa Delta Phi Record*, December, 1976, 40-42.

Spore, L. The competencies of secondary school science teachers. *Science Education*, *46* (4), 1962, 319-334.

Stolovitch, H. D. Future forecasting in teacher education *AECT Research and Theory Division Newsletter*, February, 1977, *6* (1), 13-19. Washington, D.C.: Association for Educational Communications and Technology. (Eric Document Reproduction Service No. ED 136-823.)

THE INFLUENCE OF CLASS ABILITY LEVEL ON BIOLOGY TEACHING PRACTICES

JULIE P. SANFORD
RESEARCH AND DEVELOPMENT CENTER
FOR TEACHER EDUCATION
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS

INTRODUCTION

Recent research on teaching reflects increasing attention to context variables such as class composition and, to a lesser extent, to curriculum issues. For a discussion of both trends see Good (1982) and Doyle (in press). Several recent studies have suggested that class ability levels (e.g., mean student aptitude and its variance within a class) have significant effects on classroom teaching processes, classroom management, and student achievement gains. Examples of such studies include Beckerman and Good (1981), Evertson (1982), Evertson, Sanford, and Emmer (1981), and Veldman and Sanford (1982). None of these studies were conducted in the subject area of secondary science. The present study examines the effect of class ability level on teachers' use of inquiry teaching strategies in high school biology classes.

BACKGROUND

During the 1960's and 1970's, secondary science education emphasized an inquiry approach to teaching and learning in the classroom. Current Biological Sciences Curriculum Study (BSCS) textbooks and other modern science curriculum materials retain some emphasis on inquiry teaching methods. The inquiry approach may be defined as a set of teaching strategies that present science as a human activity as much as a body of knowledge and that emphasizes development of science process skills in students (Lee, 1961; Schwab, 1962). Examples of students' cognitive processes emphasized in the inquiry approach to biology teaching include formulating questions, selecting information pertinent to questions, making inferences, applying rules and concepts, organizing information, formulating hypotheses, designing simple experiments, and interpreting data (Bingman, 1969).

Descriptive studies of instruction in secondary science classes have indicated relatively little use of inquiry activities or process-oriented approaches to the curriculum (Hurd, Bybee, Kahle, & Yager, 1980; Stake & Easley, 1978). Some teachers appear to assume that the academic ability level of students in their biology classes affects the extent to which they as teachers can and/or should use inquiry teaching approaches (Hurd, et al., 1980). This widely held notion might appear to be supported by some research findings that show that students' ability to perform some specific inquiry tasks is directly related to students' academic aptitude (Cheong, 1971; Gibbs, 1967; Raun & Butts, 1967). These and others have reported relationships between individual student performance on certain inquiry tasks (such as formulating a hypothesis, problem solving, or designing a study) and academic aptitude. In addition, results of studies by Evertson (1982), Veldman and Sanford (1982), and Metz (1978) suggest that classroom management problems associated with lower ability classes might affect the successful use of some inquiry activities in lower ability classes (e.g., frequent class discussions, laboratory investigations).

Whether student or class success in inquiry activities is related to student academic ability, there are no grounds for assuming that low ability students should not be given the opportunity to practice inquiry in their science classes. Indeed, the need for such practice might be indicated. Studies by Gibbs (1967), Raun & Butts (1967), Milson (1972), and others have shown that students of many ability levels can benefit and improve specific skills with inquiry training or practice. Lee (1961), reported that in the testing program for four of the BSCS Lab Blocks, classes of many ability levels experienced success with the inquiry-oriented laboratory-centered materials.

Classroom interaction research by Lundgren (1972) in another content area demonstrated significant relationships between class ability levels and occurrences of many behaviors (both teacher and student) pertinent to inquiry in classrooms. The most significant correlations between classroom behaviors and student ability levels resulted when correlations were computed using the average IQ of the "steering group" as the ability variable. Lundgren defined the steering group of each class as those students whose IQ scores fell between the 10th and the 25th percentiles for their class. He asserted that teacher and student behaviors were less a function of the mean ability levels of classes than of the average intelligence of a relatively small group of students near the bottom of the ability continuum of each class.

Results of several studies suggest indirectly that inquiry strategies may in fact be used more with higher ability classes than with lower ability classes. Research by Hillis (1975), Patelmo (1975), and Jingozian (1973) have been discussed in some detail by Sanford (1977). These studies made use of the *Biology Classroom Activity Checklist* (BCAC) or some modification of the BCAC. This instrument developed by Leonard Kochendorfer (1967) measures classroom use of inquiry teaching strategies, using student perceptions of classroom activities. The Kochendorfer instrument was developed to measure congruency with the philosophy of a particular curriculum project, the BSCS, and as such the composite score it yields has limited application. However, most secondary school science programs include among their curriculum goals the development of some science process or general thinking skills; the BCAC, with seven subscores assessing different aspects of classroom activities, has potential for providing information about the use of teaching activities related to these goals. On a more general level, it may also be used to study the question of how teachers vary their teaching strategies in different classroom contexts such as class composition.

PROBLEM

The present study investigated the use of inquiry teaching strategies as a function of the academic aptitude of students in high school biology classes.

The study attempted to answer the following research questions:

1. Is there a significant relationship between class academic aptitude as measured by the *Short Form Test of Academic Aptitude* and the extensive use of inquiry strategies in high school biology classes, as measured by the *Biology Classroom Activity Checklist* (BCAC)?
2. To what extent does individual teachers' use of inquiry strategies vary from class to class, and when higher and lower mean aptitude classes for each teacher are compared?
3. To what extent are individual students' *perceptions* of classroom teaching practices on the BCAC related to their individual academic aptitude? (This question challenged the assumption made in questions 1 and 2 that mean class BCAC scores are objective measures of actual class practices, and investigated the extent to which individual students' responses on the BCAC instrument are related to the students' academic aptitude. Investigation of this question served as an indirect check on the validity of the BCAC, and results could have implications for the use of student perception instruments in general.)

RESEARCH DESIGN

The research design used to address these questions was that of a descriptive study including the following three separate analytical strategies: correlations of mean class BCAC scores (including seven subscores) with two measures of class academic aptitude (Research Question 1); analysis of variance in teaching practices comparing mean BCAC scores for different classes taught by the same teacher (Research Question 2); and multiple regression of individual students' aptitude scores and individual students' BCAC ratings of their class, controlling statistically for differences in means BCAC scores between classes (Research Question 3).

INSTRUMENTATION

The Biology Classroom Activity Checklist (BCAC) consists of statements describing activities in the classroom, to which students respond using a true/false format. The mean of students' scores in a particular class is used as an indicator of the extent to which inquiry teaching strategies are used in that class. BCAC statements are arranged in seven sections, each focusing on a different aspect of class activities or teaching strategies and yielding a separate subscore. In general, instrument items that contribute positively toward inquiry objectives describe an emphasis on higher-order questions and tasks, less teacher emphasis on lecture and the textbook as sources of all knowledge, frequent use of laboratory to investigate problems related to other class work, pre-lab discussion of the purpose of activities but not the answers to the lab problems, teacher supervision and questioning of students during lab, some opportunity for individual student investigation, and use of lab for at least one-fourth of the biology class time.

The reliability coefficient of the biology classroom activity checklist was reported by Kochendorfer (1967) at .96, based upon the Horst reliability measure comparing intra-class and inter-class variances among 64 classes studied in 1966.

The *Short Form Test of Academic Aptitude* instrument is a group administered written test yielding an IQ. It is a revision of the widely used *California Short Form Test of Mental Maturity.* For more information about these two instruments see Goldman (1972).

POPULATION

In 59 first year biology classes, 1,148 students taught by 15 different teachers comprised the same for the study. All biology teachers in three high schools in a southwestern city of 116,200 participated. The three schools represented a wide range of socioeconomic groups within a diverse community. Classes were somewhat homogeneously grouped according to school achievement test scores, with the sample consisting of 46 regular classes, eight basic classes, and five accelerated classes.

All classes were following the same basic first year biology curriculum, using *Biology* by Smallwood and Green (1972) as a textbook, with the exception of the accelerated classes, which used *Biological Science, An Inquiry Into Life* by the Biological Sciences Curriculum Study (1963).

DATA COLLECTION

The BCAC was administered by the teachers in all of their first year biology classes during two different trimesters. The number of classes per teacher in the total sample varied from one to seven. Data from classes where fewer than half of the students participated were discarded. Student participants' scores on the *Short Form Text of Academic Aptitude* were obtained from central administration records.

RESULTS

Table 1 shows class mean aptitude scores and mean BCAC scores for all classes by assigned ability level. The average class mean IQ for the entire sample was 99.5, while basic classes averaged 82.3 and accelerated classes averaged 117.6. The range of class mean IQ scores within the entire sample was 77.6 to 121. Within the 46 regular classes, mean class scores ranged from 91.2 to 106.6. The average IQ of the steering groups (made up of students whose IQs fell in the 10th to the 25th percentiles for their classes) were approximately 10 points below class mean IQ scores.

Table 1
CLASS MEAN ABILITY LEVELS AND BCAC SCORES

	All Classes	**Basic**	**Accelerated**	**Regular**
Number of Classes	59	8	5	46
Mean IQ				
Class	99.5	82.3	117.6	110.48
Steering Group	89.4	72.9	106.6	90.4
BCAC Total Score	49.4	51.5	56.8	49.5
BCAC Section Scores				
A: Teacher Role	55.6	52.3	61.0	56.1
B: Student Role	52.6	55.4	60.4	52.2
C: Text Use	42.2	47.9	51.4	42.5
D: Tests	37.1	42.7	69.7	33.7
E: Lab. Prep.	58.2	54.7	62.9	58.3
F: Laboratory	50.4	53.1	49.5	50.6
G: Post Lab.	51.8	51.9	51.7	50.8

Class mean BCAC scores ranged from 40.0 to 65.2 for the total or composite score (possible range zero to 100). Scores differed widely from subsection to subsection of the instrument. For the entire sample of 59 classes, the class mean scores were, on the average, lowest for sections D, Design and Use of Tests, and C, Use of Curriculum Materials, and highest for section E, Lab Preparation, and section A, Role of the Teacher.

The reliability coefficient obtained using the Horst formula for the sample for this study was .92, only slightly lower than that of the .96 reported by Kochendorfer (1967).

RESEARCH QUESTION 1

Research Question 1 examined relationships between measures of class academic aptitude and the extent of use of inquiry teaching strategies. Table 2 shows the Pearson correlations between class mean IQ and BCAC scores for the 59 classes. There were statistically significant positive correlations between mean class academic ability and mean BCAC scores pertaining to the role of the teacher, the use and design of tests, and laboratory preparation. The IQ correlation with section A, Role of the Teacher, was strongest and significant at the .01 level. There was a statistically significant negative correlation between class mean academic ability and the BCAC section dealing with laboratory activities. If mean class BCAC subscores are accepted as indicatoros of actual classroom practices, then it appears that teachers tend to be more inquiry-oriented with higher aptitude classes with respect to teacher role, testing, and laboratory preparation, but less inquiry-oriented with respect to laboratory activities. (An alternative hypothesis is that higher aptitude classes perceive some class practices differently than do lower aptitude classes. The alternative hypothesis was tested under Question 3.)

Table 2

PEARSON CORRELATION COEFFICIENTS FOR MEASURES OF CLASS ABILITY LEVEL AND CLASS MEAN BCAC SCORES

BCAC Subscores and Total Scores	Correlation with Class Mean I.Q.	Correlation with Steering Group Mean I.Q.
Total BCAC Score	.0182	−.0479
BCAC Subscores*		
Section A: Role of the Teacher	.3015	.2016
Section B: Student Participation	−.0042	−.0783
Section C: Curriculum Materials	−.1277	−.0997
Section D: Tests	.2350	.0860
Section E: Laboratory Preparation	.2245	.1698
Section F: Laboratory Activities	−.2529	−.2077
Section G: Laboratory Followup	−.1503	−.0671

*Underscored coefficients indicate correlations which meet or exceed the .05 level of significance.

In general, however, results obtained in investigating Question 1 indicated very little direct relationship between class academic aptitude levels and *overall* extent of use of inquiry strategies as indicated by the BCAC *total* score. On the total score and four of the subscores, regular classes scored *lower* than either the basic or accelerated classes. Although accelerated classes' average mean score exceeded that of the basic and regular classes on the total BCAC and every subscore but F, Lab Activities, sample limitations allowed few conclusions to be made about these findings (see Discussion section).

To test Lundren's (1972) theory about the effects of steering group ability level on teacher behavior, a similar set of correlations were computed between this measure of class composition and class BCAC scores. These correlations are also included in Table 2. None of the correlation coefficients were significant at the .05 level. The steering group mean IQ correlated highly with the class mean IQ scores, but consistently yielded less significant correlations with inquiry teaching practices (BCAC scores) than did class mean IQ. To the extent that inquiry teaching practices appeared to be related to class ability levels, the class average aptitude appeared to be a more closely related factor than that of the student group identified by Lundgren.

RESEARCH QUESTION 2

Question 2 was "To what extent does individual teachers' use of inquiry strategies vary from class to class, and when high and lower mean aptitude classes for each teacher are compared?" Hypotheses were formulated which required three series of analysis of variance: An *F* test between BCAC class mean scores for all classes of each teacher, a correlated samples *t*-test between BCAC mean scores for the highest mean IQ class and the lowest mean IQ class for each and all teachers, and a correlated samples *T*-test between mean BCAC scores for the highest aptitude class and the lowest aptitude class of only those teachers having assigned ability level split loads (and consequently two classes of widely differing mean academic aptitudes). Analysis of variance among the class mean BCAC scores for all classes of each of the 15 teachers, assessing the consistency with which teachers use inquiry strategies with all their classes indicated significant differences among classes of Teachers 09, 10, and 11. The remaining 10 teachers did not vary significantly (at the .05 level) from class to class with respect to the inquiry teaching strategies measured by the BCAC. In the cases of Teachers 09 and 11, the variance in BCAC mean scores was clearly *not* related to variance in class mean IQ. Among the classes of Teacher 10, however, variance in class mean BCAC did appear to be roughly related to variance in mean class aptitude scores.

Analysis of variance computed on the mean scores for *different* teachers showed that BCAC class mean scores varied significantly ($p < .001$) from teacher to teacher. This fact, combined with the low or not significant correlations obtained for class mean total BCAC scores with class mean aptitude (under Question 1) indicated that *teacher characteristics* were much more significant determiners of classroom inquiry practices than were ability levels of students in the classes.

To compare teaching strategies used in teachers' highest and lowest classes, a series of correlated samples *t*-tests were carried out to compare class mean BCAC scores for 13 highest aptitude classes were compared with those of 13 lowest aptitude classes. The highest aptitude and lowest aptitude groups of classes were each composed of one class from each of the 13 teachers having more than one class. Included were seven teachers having split loads (classes labeled by the district as regular and basic or regular and accelerated). The test assessed the extent to which the 13 teachers, as a group, used different strategies with their highest aptitude classes compared with their lowest aptitude classes. Results of the *t*-tests of the seven BCAC subscores and total showed that only one subscore, Section A (Role of the Teacher), differed significantly between teachers' highest IQ classes and lowest IQ classes. The mean score for the highest aptitude classes was 60.44, and the mean score of the lowest classes was 50.18 ($t = 6.23$; $p<.001$). With regard to the BCAC total score and 6 other subscores, no significant differences were found. Results were similar when only the seven teachers with split loads were considered.

RESEARCH QUESTION 3

The final research question in the study asked: "To what extent are individual students' perceptions of classroom teaching practices on the BCAC related to their individual academic aptitudes?" To answer the question, multiple regression analysis was used to assess the extent to which students' IQs and class membership were a better predictor of individual BCAC scores than was class membership alone. Table 3 shows for each BCAC section and total BCAC:

1. The R^2 computed between class membership and individual BCAC scores (the restricted model),
2. The R^2 computed between class membership and IQ scores together and the BCAC scores (the full model),
3. The value of *F* computed on the difference between the two R^2, and
4. Indication of significance at the .05 level or better.

Data in Table 3 indicate that for the BCAC total score and subscores B, D, and E, pupil IQ was not a significant factor accounting for variance in student perception of classroom activities. On sections A, C, F, and G of the instrument, however, individual IQ appeared to significantly affect scores the students gave their teachers on the BCAC. Thus, there clearly were some pupil IQ/pupil perception interaction effects influencing BCAC scores with respect to items on the role of the teacher, use of text and curriculum materials, laboratory activities, and lab followup. The magnitude of IQ/perception interaction effects for the subgroups was not large, however, and effects on the BCAC total score were not significant.

DISCUSSION

The purpose of this study was to assess the effects of class composition as a context variable affecting teaching strategies in biology classes. The investigation also provided opportunity to learn more about the use of pupil perceptions as sources of information about classroom practices. Results of analyses for questions 1 and 2 indicated that variation in class mean BCAC (total) scores was not significantly affected by class ability level, whether the indicator of class ability level was the class mean IQ score or the mean scores of class steering groups. Most of the teachers in this sample were consistent

Table 3

RESULTS OF MULTIPLE REGRESSION OF INDIVIDUAL STUDENTS' IQ AS A FACTOR RELATED TO BCAC SCORES (STUDENTS' PERCEPTION OF TEACHING STRATEGIES)

	R^2 Restricted	R^2 Full	Value of F[1]
BCAC Total Score	.28819	.28844	0.3826
BCAC Section Scores			
A: Role of the Teacher	.16345	.16807	6.0420*
B: Student Participation	.18583	.18597	0.1871
C: Curriculum Materials	.21332	.21909	8.0395**
D: Tests	.34517	.34688	2.8490
E: Laboratory Preparation	.10415	.10535	1.4593
F: Laboratory Activities	.19492	.20311	11.1819
G: Laboratory Followup	.12226	.12761	6.6772

[1] df for all F tests = 1, 1088.

*A single underscore indicates significance at $p \leq .05$.

**A double underscore indicates significance at $p \leq .01$.

in their use of teaching practices, as indicated by the mean BCAC total scores for each of their classes. With the exception of one teacher, they were as inquiry-oriented with their lower aptitude classes as with their higher aptitude classes. Considerable variation was found across teachers, however, indicating that *teacher characteristics* are more significant determiners of classroom inquiry practices than is class composition.

Comparison of BCAC scores for different assigned ability levels (regular, basic, and accelerated classes) suggests two comments. First, basic classes tended to be, if anything, more inquiry-oriented than regular classes, not less. Second, a serious limitation of the present sample prevented us from drawing conclusions about the accelerated classes, although mean BCAC scores for these classes were higher than other classes (as might be expected considering that these classes were the only ones using a BSCS textbook). Unfortunately, there were only five accelerated classes in our sample and four were taught by one teacher. The teacher who taught four accelerated classes also taught three regular classes (in two trimesters), and the BCAC scores in her regular classes were not significantly different from those in her accelerated classes. This supports the generalization that the use of inquiry strategies is related more to teacher characteristics than to class composition (or textbook use). The other teacher of an accelerated class had higher BCAC scores in her accelerated class than in her regular classes. Nevertheless, using the methods of the present study, a larger sample of accelerated classes would be needed to draw any conclusions about classroom practices in accelerated classes.

Considerable variation was found in class scores across the seven subsections of the BCAC. The section scores also varied some in their relations to class and individual student ability levels. As a whole classes in the sample were least inquiry-oriented with regard to use of curriculum materials and testing practices. Scores for these key components of science instruction suggest for this sample a heavy dependence on the textbook, little use of other reference materials, and frequent use of tasks such as answering review questions and outlining the text. Scores for testing practices suggest extensive use of lower order questions. These findings are consistent with findings of recent surveys of science education practices (Hurd, et al., 1980; Stake and Easley, 1978).

Class mean scores for the testing section showed great variation across assigned ability levels, and they correlated significantly with class mean aptitude scores in the whole sample. Considering that the testing subscore was *not* found to be significantly affected by student perception differences related to ability, results suggest that teachers are more likely to adjust *testing practices* to class aptitude levels than they are other classroom practices. The only other aspect of classroom practices in which significant ability level correlations and no significant student ability perception differences were found was laboratory preparation, suggesting that teachers are more likely to describe laboratory objectives (but not answers) and to link laboratory meaningfully to other class work when they are teaching higher or average ability classes than when teaching lower ability classes. It should be noted, however, that although ability level correlations for both testing practices and pre-lab activities were significant at the .05 level, both were low (r = .24 and .22), indicating that 5% or less of the variation in class BCAC scores could be accounted for by variation in class mean IQ.

Findings that for the most part teachers do not vary their classroom practices according to the ability levels of classes have implications for school ability grouping practices, at least to the extent that homogeneous grouping is done on the assumption that the practice will allow teachers to use teaching strategies they consider most appropriate for each group of students. Apparently little such variation occurs with respect to most classroom practices on the BCAC. Evertson (1982) has also reported that in English and mathematics classes in junior high school teachers used similar class activities with classes of different ability, although activities were more successfully and smoothly conducted in higher ability classes than in lower ability classes.

Regression analysis revealed that the BCAC (total) score is not significantly affected by variation in student perception according to individual academic aptitude levels. However, findings indicated that for four of the seven BCAC subscores pupil I.Q. was a significant factor accounting for variance in individual scores. Although these effects were not large, findings of student I.Q./student perception effects influencing scores on some sections of the BCAC tempered interpretation of low, but significant correlations between mean class aptitude and mean BCAC scores for two sections of the BCAC: Role of the Teacher (section A) and Laboratory Activities (section F). Higher ability students tended to perceive the roles of their teachers as more inquiry-oriented than lower ability students did (even in the same class), but they perceived laboratory activities to be less inquiry-oriented. Examination of items in section A, Role of the Teacher, reveals that six of the seven items describe verbal interactions that may in fact vary for individual students in a class. The BCAC assumes that student responses pertain to class and teacher activities as a whole, but students' responses very likely are affected by individual interactions with the teacher. For example, the statements "my teacher often asks us to explain the meaning of certain things in the text," and "my teacher often asks questions that cause us to think about the evidence that is behind statements that are made in the textbook," may not describe teachers' practices with students in general, but may in fact describe their interactions with the higher ability students in a class. These students would respond differently on these items than would lower ability students. Items in the section on laboratory activities do not appear to lend themselves to a similar explanation of student perception differences related to ability. It may be that higher ability students are more likely to realize the limitations of their laboratory experiences than lower ability students are. Instrument sections on curriculum material and lab followup also appeared to be affected by student perception differences, but no correlations with class ability level were found for these subscores in this sample. Student perception differences for four subscores of the BCAC should be considered in interpreting the results of any study using the instrument, particularly when comparing classes of very different ability levels. In general, however, variations in pupil academic aptitude should not be considered a significant threat to interobserver reliability (and thus, to the validity of the instrument as an indicator of practices in science classrooms) when mean class BCAC total scores are used. This study confirmed the high reliability coefficient reported by the developer of the instrument (Kochendorfer, 1967).

It is recognized that any instrument that uses pupils' perceptions to assess teaching practices has many limitations with respect to validity. The ideal way to establish the validity of the BCAC would be to compare results using it in several classes with direct observation of those classes. Examination of the instrument shows that many of the variables on the instrument would be very difficult to assess reliably through a limited number of classroom observations, however. The BCAC and similar pupil perception instruments that could be developed have potential value as unobtrusive, inexpensive, and easily administered measures of long-term, normal classroom practices and environment. Continued use of the BCAC in research, or more importantly, its use as a prototype for the development of similar instruments using pupil perceptions for research in teaching, appears to be warranted.

REFERENCES

Beckerman, T., & Good, T. The classroom ratio of high- and low-aptitude students and its effect on achievement. *American Educational Research Journal*, 1981, *18*. 317-327.

Bingman, R. M. (Ed.) Inquiry Objectives in the Teaching of Biology Midcontinental Regional Educational Laboratory and Biological Sciences Curriculum Study. *Position Paper 1* (1), 1969.

Biological Sciences Curriculum Study. *Biological science, an inquiry into life*, 2nd edition. New York: Harcourt, Brace and World, Inc., 1963.

Cheong, S. Y. *An analysis of inquiry performance of high school biology students including the relationship of inquiry performance to instructional technique and to student achievement and academic ability.* Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1971.

Doyle, W. Paradigms for research on teacher effectiveness. In L. S. Shulman (Ed.), *Review of research in education (Vol. 5)*. Itasca, Illinois: F. E. Peacock, 1977.

Evertson, C. Differences in instructional activities in higher- and lower-achieving junior high English and math classes. *Elementary School Journal*, 1982, *82*, 329-350.

Evertson, C., Sanford, J., & Emmer, E. Effects of class heterogeneity in junior high school. *American Educational Research Journal*, 1981, *18*(2), 219-232.

Gibbs, R. K. *An analysis of the effectiveness of the BSCS single topic films in teaching hypothesis construction in high school biology students*. Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana, 1967.

Goldman, B. A. Review of California Test of Mental Maturity. In O. K. Buros (Ed.), *The Second Mental Measurements Yearbook*. Highland Park, New Jersey: Gryphon Press, 1972.

Good, T. *Classroom research: What we know and what we need to know*. Austin: Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, 1982.

Hillis, S. R. *The relationship of inquiry orientation in secondary physical science classrooms to students' critical thinking skills, attitudes, and views of science*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1975.

Hurd, P. D., Bybee, R. W., Kahle, J. B., & Yager, R. E. Biology education in secondary schools of the United States. *The American Biology Teacher*, 1980, *42*(7), 388-410.

Jingozian, C. R. *A study of the relationship between certain teacher practices and student attitudes in the secondary science classroom*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1973.

Kochendorfer, L. H. The development of a student checklist to determine classroom teaching practices in high school biology. In A. E. Lee (Ed.), *Research and Curriculum Development in Science Education* (The University of Texas Publication #6720). Austin, Texas: Science Education Center, The University of Texas at Austin, 1967.

Kochendorfer, L. H. Classroom practices of high school biology teachers using different curriculum materials. In A. E. Lee (Ed.), *Research and Curriculum Development in Science Education* (The University of Texas Publication #6720). Austin, Texas: Science Education Center, The University of Texas at Austin, 1967.

Lee, A. E. Biology Laboratory instruction innovation. *The Science Teacher*, 1961, *28* (6).

Lee, A. E. The experimental approach in teaching biology. *American Biology Teacher*, 1961, *23*(7).

Lundgren, U. P. *Frame factors and the teaching process*. Stockholm: Elmquist and Wiksell, 1972.

Metz, M. H. *Teachers' adjustment to students' behavior: Some implications for the process of desegregation*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March, 1978.

Milson, J. L. The development and evaluation of physical science curriculum materials designed to improve student attitudes. *Journal of Research in Science Teaching*, 1972, *9*(4).

Patelmo, P. J. *The relationship of selected teacher characteristics and classroom behaviors of teachers to critical thinking growth in students using the Introductory Physical Science (IPS) program as the content vehicle*. Unpublished doctoral dissertation, Temple University, Philadelphia, Pennsylvania, 1975.

Raun, C. E., & Butts, D. P. The relationship between the strategies of inquiry in science and student cognitive and affective behavioral change. *Journal of Research in Science Teaching*, 1967-1968, *5*.

Sanford, J. P. *Class academic aptitude level as a factor affecting the use in inquiry strategies in high school biology classes*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1977.

Schwab, J. J. The teaching of science as inquiry. In J. J. Schwab and P. F. Brandwein (Eds.), *The Teaching of Science*. Cambridge, Mass: Harvard University Press, 1962.

Smallwood, W. L. & Green, E. R. *Biology*. Morristown, New Jersey: Silver Burdette Company, 1972.

Stake, R. E., & Easley, J. A. *Case studies in science education* (Volumes I & II). Urbana, Illinois: Center for Instructional Research and Curriculum Evaluation and Committee on Culture and Cognition, The University of Illinois at Urbana-Champagne, 1978.

Veldman, D. J., & Sanford, J. P. *The influence of class ability level on student achievement and classroom behavior* (R&D Rep. No. 4095). Austin, Texas: Research and Development Center for Teacher Education, The University of Texas at Austin, 1982.

EVALUATION OF AN ENVIRONMENTAL SCIENCE EDUCATION ACADEMIC YEAR PROGRAM FOR INTERMEDIATE ELEMENTARY TEACHERS (4-6)[1]

LOWELL J. BETHEL
SCIENCE EDUCATION CENTER
COLLEGE OF EDUCATION
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS

SHIRLEY M. HORD
RESEARCH AND DEVELOPMENT CENTER
FOR TEACHER EDUCATION
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS

INTRODUCTION

Inservice education has been roundly criticized for many years (Bethel & Hord, 1981). It has failed inservice teachers because it rarely has met their needs. The inservice literature reports that it has been held at inappropriate times, provided little useful information or ideas, and is perceived as just plain dull (Bailey & James, 1978; Lawrence, 1974; Seldin, 1979). Many efforts have been made to provide relevant and meaningful inservice but few have reported any reliable research or data designed to substantiate it.

Many inservice teachers desire to continue their professional training and development beyond the normal four year teacher preparation program. Nowhere is this more evident than at the elementary school level.

An area in which inservice teachers look for training is science (Bethel, 1982). Elementary school teachers are generally uncomfortable in teaching science because often they do not have a sufficient knowledge base or understanding of the discipline. Thus they tend to ignore or give little time to science instruction. This set of circumstances has led to an almost total abandonment of science instruction at the elementary school level (Bethel, 1980).

The National Science Foundation (NSF) has as one of its primary objectives the improvement of science, taught in grades K-12. In order to carry out this mandate the Pre-College Teacher Development in Science Program was initiated in 1959. Federal monies allocated to the NSF were distributed nationwide to both colleges and universities in order to develop meaningful and effective science inservice programs for teachers.

Such an effort was undertaken recently at the Science Education Center located in Austin at The University of Texas (Bethel, 1981). One of it's goals has been to provide assistance and support for teachers already in service. The Center has been involved in providing staff development opportunities for teachers from south central Texas since 1960.

---

[1] This material is based upon work supported by the National Science Foundation and the National Institute of Education. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institute of Education.

NEEDS ASSESSMENT

Before applying for NSF funds, personnel from the Center visited schools and observed what was being taught. It was found after extensive observations that little or no science was being taught (for a detailed description see Bethel & Hord, 1981). The observations were followed up by a survey questionnaire which asked elementary school teachers what kind of inservice they would like to see and what areas needed to be addressed. The overwhelming response was science and in particular environmental science (Bethel, 1982). Based on input from inservice teachers, administrators, curriculum coordinators, and other professionals a grant was submitted to NSF and funded. The group identified as most interested in an environmental science education program was intermediate elementary inservice teachers in grades 4-6. The grant provided for one year in order to work with 48 teachers.

The remainder of this paper will be a description of the environmental science education inservice program. The research methods and techniques undertaken to study change in the program participants, will be reported also. Then the results and findings of the inservice research study will be presented. Finally, implications of the findings for science inservice will be suggested.

INSERVICE PROGRAM

Based upon the objectives of NSF and the needs expressed in the survey, the following objectives were identified for the environmental science education program:

1. To improve the knowledge of intermediate elementary teachers in grades 4-6 from the central Texas region in environmental science.
2. To aid teachers in the identification and use of environmental science education resources appropriate to their level of instruction which would aid in the teaching of environmental science concepts.
3. To develop and maintain cooperation, communication, and program support between scientists at The University of Texas at Austin and intermediate elementary teachers in the central Texas region.

The needs assessment revealed that teachers were very much interested in environment-related concepts. They expressed interest in such topics as ecosystems, environments, minerals and rocks, air pollution, water pollution, communities, energy, and conservation. Therefore, these topics were included in the course content along with environmental law, geology of the Southwest and land management.

A total of 32 class sessions were held during the year. Participants attended all classes, each of which were 2½ hours long. The lectures, demonstrations and seminars usually required 75 minutes. The laboratory sessions lasted for

60 minutes followed by a 15-20 minute post-lab discussion to bring closure to the session and to clarify questions raised by the participants.

The materials and learning activities utilized in the class sessions were developed by the project director, inservice teachers, Texas Education Agency personnel, and members of the university's multidisciplinary faculty who served as resource personnel, including:

1. a biologist
2. a geologist
3. a botonist
4. an environmental chemist
5. an ecologist
6. a science educator
7. an environmental geologist

To enhance the instructional program, university scientists having expertise in many of the program topics were invited to present mini-lectures for the participants. The purpose for having the scientists was two-fo[illegible] (1) to present the latest in environmental science knowledge and technology, and (2) to introduce university scientists who could serve as future resource persons to the participants and their environmental science education programs conducted in the schools.

In order to provide program activities and materials that would be more relevant for the teachers, data, which were collected as a measure of change in the participants for summative evaluation purposes, were also used as formative feedback to the program director. The data made it possible to better understand and respond to participants' needs as they experienced the inservice change process. Thus, program modifications were designed and targeted toward the changing concerns and needs of the teachers.

A unique aspect of the program was the conducting of inservice workshops by the participants for other teachers in their respective schools and school systems. This was referred to as the "multiplier" effect. The overall purpose was for the participants to share their knowledge and expertise not only with their pupils, but with their colleagues and, ultimately, colleagues' students. In this manner the cost-effectiveness of the NSF environmental science education program would be enhanced in terms of the number of people involved with the total program. The number of workshops given by the participants was 35 which were attended by an average of 25 inservice teachers each, reaching 875 teachers. If each teacher worked with an average of 27 students, then a total of approximately 23,625 students would have been exposed to science instruction and in particular, environmental science education.

PARTICIPANTS

In order to secure applicants, program notices were sent to neighboring school districts within 75 miles of the university. In addition to direct mailings, information was distributed through state science teachers' newsletters, the Texas Education Agency, and regional education service centers.

Applicants expressing an interest in the program were required to return an application and two letters of support — one from an administrator supporting the teacher and one from a teaching colleague. Teachers were selected on the basis of the following criteria:

1. Receipt of application and supporting letters of recommendation
2. At least two years of teaching experience in grades 4-6
3. Selection without reference to sex, race, creed, national origin
4. Minimum of 3 to 6 hours of undergraduate science courses
5. Teaching assignment in intermediate, elementary or middle school (grades 4-6)

Forty-eight inservice teachers were chosen from the central Texas region to participate in the environmental science education program. The final group of applicants chosen had an average of 4.6 years of teaching experience, an average age of 27.5 years, and an average of 5.1 semester hours of undergraduate science.

RESEARCH METHODS AND TECHNIQUES

In order to determine the effects of the inservice program on the participants and to realize a measure of its success, a study was designed to obtain formative and summative evaluation data from program participants. In this way, results from ongoing data collection could be used as a diagnostic basis for instituting changes, thus insuring the refinement of the program to meet the immediate needs and concerns of the teachers.

INSTRUMENTATION

Two instruments were selected to provide data. In order to determine teachers' concerns about the innovative science program and how concerns might change as a result of the inservice program, the Stages of Concern Questionnaire (SoCQ) was selected. Second, to measure teachers' attiudes toward science and how they might be affected by the inservice intervention, the Environmental Education Questionnaire (EEQ) was chosen. The underlying concepts and assessment tools for these two measures are described below.

Stages of Concern Questionnaire. As a result of four years of engaging in research on change in schools and colleges, the Concerns-Based Adoption Model (CBAM), an important conceptual framework useful in understanding, monitoring and facilitating change in individual teachers was developed at the Research and Development Center for Teacher Education at The University of Texas at Austin. The CBAM offers a unique approach to the study of teacher change by focusing on the needs and describing the growth of individuals over time; it describes teachers as they first begin and then gain more experience with educational innovations. One of the dimensions of this model is Stages of Concern (SoC) (Hall, George & Rutherford, 1977), a conceptualization of the way the *concerns* of individual teachers change as they become familiar with and involved with new programs, processes or educational practices in their schools.

By *concern* is meant the feelings, attitudes, thoughts, ideas, or reactions an individual has related to a new practice. The work of Frances Fuller (1969) focused on the concerns of teachers-in-training as they progressed from early experiences in preservice teacher education programs to being experienced inservice teachers. This sequence of teacher concerns Fuller labelled as *unrelated, self, task, impact*. Fuller's work was the base upon which CBAM research on Stages of Concern was built. A result of this research is the reaffirmation that not only do teachers go through a sequence of concerns about teaching, but all teachers faced with a new program or innovation have concerns that are identifiable and developmental and are similar to those documented by Fuller. From this research on change, seven *Stages of Concern About the Innovation* have been identified (Figure 1).

Figure 1

STAGES OF CONCERN ABOUT THE INNOVATION*

	STAGES OF CONCERN	DEFINITIONS
IMPACT	6 REFOCUSING	The focus is on exploration of more universal benefits from the innovation, including the possibility of major changes or replacement with a more powerful alternative. Individual has definite ideas about alternatives to the proposed or existing form of the innovation.
	5 COLLABORATION	The focus is on coordination and cooperation with others regarding use of the innovation.
	4 CONSEQUENCE	Attention focuses on impact of the innovation on student in his/her immediate sphere of influence. The focus is on relevance of the innovation for students, evaluation of student outcomes, including performance and competencies, and changes needed to increase student outcomes.
TASK	3 MANAGEMENT	Attention is focused on the processes and tasks of using the innovation and the best use of information and resources. Issues related to efficiency, organizing, managing, scheduling, and time demands are utmost.
SELF	2 PERSONAL	Individual is uncertain about the demands of the innovation, his/her inadequacy to meet those demands, and his/her role with the innovation. This includes analysis of his/her role in relation to the reward structure of the organization, decision making, and consideration of potential conflicts with existing structures or personal commitment. Financial or status implications of the program for self and colleagues may also be reflected.
	1 INFORMATIONAL	A general awareness of the innovation and interest in learning more detail about it is indicated. The person seems to be unworried about himself/herself in relation to the innovation. She/he is interested in substantive aspects of the innovation in a selfless manner such as general characteristics, effects, and requirements for use.
	0 AWARENESS	Little concern about or involvement with the innovation is indicated.

*Original concept from G. E. Hall, R. C. Wallace, Jr., & W. A. Dossett, A Developmental Conceptualization of the Adoption Process within Educational Institutions (Austin, Tex.: Research and Development Center for Teacaher Education, The University of Texas, 1973).

Stages of Concern About the Innovation (SoC), describes the kinds of concerns which the individual may experience across time, related to the innovation. They range from initial *self* concerns (Stages 1&2), "In what ways will I be affected by this innovation?"; to concerns related to *task* (Stage 3), "How can I make this innovation work?"; and then to concerns for *impact* (Stages 4, 5, & 6), "How will using this innovation affect my students?" Individuals experience a variety of concerns at any one time; however, the degree of intensity of different concerns will vary depending on the individual's knowledge and experience.

Thus, teachers seldom have concerns at only one stage. Teacher who are non-users of an innovation typically will have concerns high on Stages 0, 1 and 2. They are more concerned about gaining information (Stage 1) or how using the innovation will affect them personally (Stage 2). As they begin to use an innovation, Stage 3 management concerns become higher and more intense. And, when teachers become experienced and skilled with an innovation, the tendency is for concerns at Stages 4, 5, and 6 to become more intense with decreases in stages 0, 1, 2, and 3 (Hall, George & Rutherford, 1977).

To measure the seven Stages of Concern, the Stage of Concern About the Innovation Questionnaire (SoCQ) (Hall, George & Rutherford, 1977) is used. This psychometrically rigorous paper and pencil measure is especially important for research and program evaluation and consists of 35 items. Teachers respond by indicating their degree of concern on a Likert scale for each of the items. Scoring these data by computer program, or manually, results in percentile scores and a profile of concerns for the individual, or for groups.

Environmental Education Questionnaire. Teachers' attitudes toward environmental science education were assessed through use of the *Environmental Education Questionnaire (EEQ)* instrument (Jaus, 1978). Attitude is defined as a predisposition to respond favorably or unfavorably toward an object, concept, or idea. In other words, a person may respond toward the concept "environmental science education" in a favorable, neutral, or unfavorable manner. The EEQ measures a person's attitude or predispositions toward environmental science education with a score between 20 and 100 (very positive). A score of 60 would indicate a neutral attitude toward environmental education. Thus movement toward the high end of the scale would indicate a more pronounced positive attitude while movement toward the low end would indicate an increasingly negative attitude.

The instrument includes a five category Likert-type scale with responses from "strongly agree" to "strongly disagree." Jaus (1978) used the test-retest method to report the reliability of the instrument. An alpha reliability of .89 was found for the combined groups on the pretest in this present study. The content validity of EEQ was established in the initial study by Jaus (1978). A total of 78 items were submitted to five science education professors and two social sciences professors. The judges were asked to rank statements, suggestions, and/or positive attitudes toward environmental science education. The original item pool (78) was reduced to those statements receiving perfect agreement among the judges. This resulted in a final instrument containing 20 items.

ASSESSMENT PROCEDURES

The inservice research study was designed so that data would be collected at three points: pre-training, mid-way (16 weeks), and post-training (32 weeks). In addition to collecting data from the sample of teachers in the treatment group, a control group was established, from whom attitude data and concerns would be obtained for comparison purposes.

In order to minimize error variance between groups related to systematic differences among individual teachers within the two samples, the control group was generated from subjects paired with the treatment group subjects. Each participant was requested to select a teaching colleague in the same grade within their school. These subjects served as the control group. Because of the similar context it was hypothesized that the control group's situation and experiences would approximate that of the participants.

RESULTS

As indicated, data were collected using the concerns (SoCQ) instrument with the treatment group, and the attitude (EEQ) instrument with both the treatment and control groups. Comparisons of the two group's attitude scores were made at the beginning of the program, at the conclusion of 16 weeks (mid-way), and at the end of 32 weeks (post-test). Comparison was also made of the treatment group's pre-, mid-, and post-SoCQ data.

There were no significant differences between the control and experimental groups on the EEQ instruments (see Table 1). Thus, both groups held similar attitudes toward environmental education and the teaching of environmental science.

Table 1

GROUP PRETEST MEANS ON THE EEQ

Control	Experimental
78.50	83.32

The EEQ was administered to both groups at the end of 16 weeks. There were significant differences between the groups (see Tables 2 & 3).

Table 2

GROUP MEANS ON THE EEQ (16th WEEK)

Control	Experimental
76.48	86.51

There was a significant difference between the two groups with the experimental group improving significantly at the end of 16 weeks in their attitudes toward environmental science education and teaching.

Table 3

ANOVA OF EEQ MEAN SCORES (16th WEEK)

Source	SS	df	MS	F
Main effect	1468.35	1	1466.35	25.61*
Residual	3324.98	58	57.33	
TOTAL	4793.33	59	81.24	

* p ≤.05

Comparison between the mean scores of the two groups at the end of 32 weeks revealed a significant difference. Again the experimental group made significant improvements in their attitudes toward environmental science and teaching (see Tables 4 & 5).

Table 4

GROUP MEANS ON THE EEQ (32nd WEEK)

Control	Experimental
78.43	88.97

Examination of Tables 1, 2, and 4 indicates that continuous improvement was made by the experimental group over the course of the environmental science inservice program.

Table 5

ANOVA OF EEQ POSTTEST MEAN SCORES (32th WEEK)

Source	SS	df	MS	F
Main effect	1341.71	1	1341.71	22.67*
Residual	2959.60	50	59.19	
TOTAL	4301.31	51	84.34	

* p ≤.05

SoCQ RESULTS

The composite profiles of the experimental group are reviewed. Next, individual profiles of some of the participants are presented and examined. The concerns of the teachers did change over the course of the environmental science education program.

The concerns of the environmental science education program participants are summarized in Figure 3. Concerns of the participants at the beginning, at the mid-point, and at the conclusion of the program were measured. Concerns in the beginning prior to the start of the program reveal a typical "nonuser" profile. Concerns were typically high on Awareness, Information, and Personal (Stages 0, 1, & 2) while low on the other Stages (3, 4, 5, & 6). The high Stage 0 might indicate that few participants were fully aware of the program and knew even less about the subject. This would probably translate as little knowledge of specific content, goals, objectives, etc. Teachers were more concerned about information than self as revealed by the higher Stage 2 and lower Stage 2. This profile indicates to the program developers an openness and desire to learn new scientific knowledge and teaching methodologies related to environmental education without undue feelings of anxiety or threat. Small peaking on Stage 5 Collaboration concerns probably reflect an interest in working with other colleagues. This is a desired characteristic since teachers in the program were required during the year to work with other teachers and to conduct a series of inservice workshops.

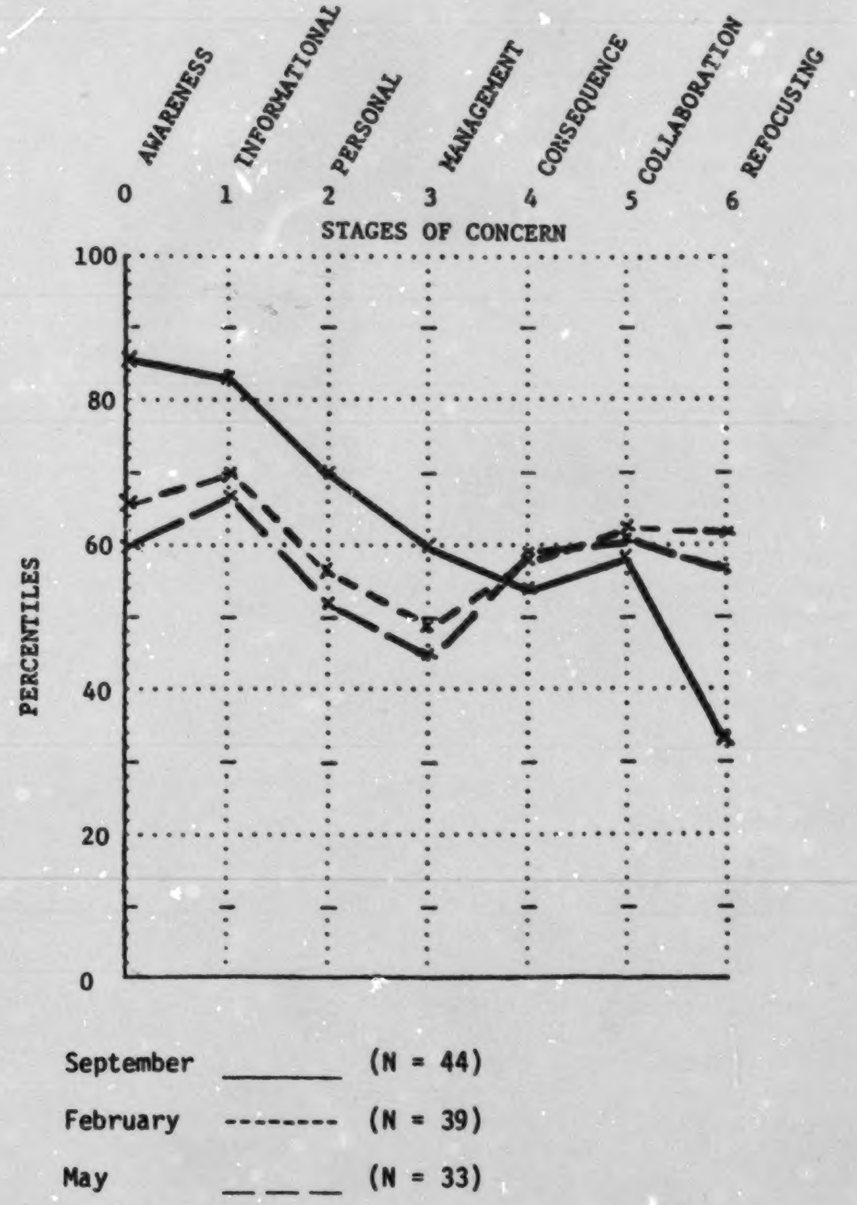
Substantial profile decreases in Stages 0, 1 & 2 continue in February and May which are the "self and task" concerns. There is a continual decrease in Management concerns indicating that participants' attention is beginning to shift probably toward Impact and Collaboration concerns.

February and May profiles are similar at both Stages 4 (Consequences) and 5 (Collaboration) and reflect a slight intensity of concerns on these "impact" stages. This is certainly a positive outcome since the environmental science program was constantly altered to meet the continuing concerns of the participants. Thus, teachers, having had a small amount of experience with teaching environmental science, would be most receptive to ideas and activities dealing with program "impact" on their students. The increased intensity of Stage 6 from September to February suggests that teachers, as a result of program experiences, had ideas about additional methods that could be incorporated to improve their science instruction.

In summary, the program results reveal that the participants' attitude toward environmental education did significantly improve during the course of the academic year. The inservice teachers' concerns changed considerably over the course of the environmental education program. These changes were viewed as useful in planning the program and specific program activities to allay their concerns. The SoCQ and EEQ proved to be useful instruments for identifying change in teachers in a specific and definite manner.

Figure 3

GROUP PROFILES OF ALL TREATMENT SUBJECTS



IMPLICATIONS AND IMPORTANCE TO INSERVICE

The data collected in this study recognized and verified the need for inservice programs which are designed to address "where teachers are," in terms of their expressed needs for teaching science and in terms of their concerns and attitudes toward new or innovative science programs. Further, the results verified the value and effectiveness of such inservice programs in enironmental science education.

In the past, the lack of appropriate diagnostic tools made the delivery of relevant inservice difficult. The instruments used and applied in this study overcame this vexing problem. In addition, the results of the study could be used as a measure of the effectiveness of an inservice program which had been implemented through utilization of quantitative data collected on teachers' concerns and attitudes during the term of the inservice program.

Change takes time. Staff developers who are responsible for inservice need to plan on a spread of training and other activities over time and paced in relationship to the concerns of teachers. A "one-shot" workshop is not likely to respond to the ongoing needs teachers have, reflected by their changing concerns as they become first knowledgeable about and then users of a program or practice. Change does take time. The more complex or complicated, the more time is required.

Change is a personal experience. Depending on many factors, the concerns of individual teachers will vary. Staff developers can use Stages of Concern to help in diagnosing, planning, delivering, and assessing activities which will be relevant and address teacher's concerns. For example, teachers with high management concerns about their science program will not respond very positively to suggestions about individualizing the lessons for increased impact on student learning.

Continuing reviews of the inservice teachers' responses on the SoC instrument were made in order to plan the materials and activities for the environmental science education program. It guided the planners in choosing teaching techniques and activities that would provide teachers with knowledge about what environmental science is, its general nature, and the major topics related to environmental science, thereby responding to their high informational concerns. Laboratory activities were designed to enhance these topics and provide teachers with "hands-on" experiences to increase their understanding. As a formative evaluation device the SoC data was very useful in monitoring the program and for making decisions about possible modifications in the program, program support and training.

One of these decisions came after reviewing the SoCQ data collected at mid-point (16 weeks). The data were studied carefully for each individual and as a result training was not conducted in a more personalized small group format (two to three participants at one time) to respond to individual's needs. In addition, emphasis was placed on activities directed to the upper levels of the SoC scale (Consequence and Collaboration Stages) as well as environmental science concepts.

During a year-long training program teachers predictably will express differential concerns and attitudes. In the example presented here, teachers did significantly change in their concerns and attitudes toward environmental science. This change continued over a long period of time (nine months). Teachers with positive attitudes toward science will tend to teach science and more than will teachers who have negative attitudes. This is an obvious statement. This study, however, shows that positive attitudes about teaching science can be significantly increased.

REFERENCES

Bailey, G. D., & James, R. K. Model for an inservice science teacher training program. *Science Education, 62* (1), 1978, 47-52.

Bethel, L. J. NSF technical report: Environmental education program results, #81-A, October, 1981, The University of Texas at Austin, Science Education Center, National Science Foundation.

Bethel, L. J. Tailoring inservice training in science to elementary teachers' needs. *Phi Delta Kappan, 63* (6), 1982, 416.

Bethel, L. J. & Hord, S. M. *A case study of change: Inservice teachers in a national science foundation environmental science education program.* Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, California, April, 1981.

Fuller, F. F. Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6* (2), 1969, 207-226.

Hall, G. E., George, A. A., & Rutherford, W. L. *Measuring stages of concern about the innovation: A manual for use of the SoC questionnaire.* Austin, Texas: Research and Development Center for Teacher Education, The University of Texas, Austin, Texas, 1977.

Jaus, H. H. The effects of environmental education instruction on teachers' attitudes towards teaching environmental education. *Science Education, 62*, (1), 1978, 79-84.

Lawrenz, F. Science teachers perceptions of their teaching skills and their school conditions. *Science Education, 58* (4), 1974, 489-96.

Seldin, C. A. Taking inservice education off the back burner. *Phi Delta Kappan*, 1979, 266, 287.

THE EFFECTIVENESS OF PUPIL FEEDBACK AND SELF-ANALYSIS IN CHANGING SELECTED SCIENCE TEACHING PRACTICES

MELANIE C. LEWIS
DEPARTMENT OF BIOLOGY
SOUTHWEST TEXAS STATE UNIVERSITY
SAN MARCOS, TEXAS

ROLLAND B. BARTHOLOMEW
SCIENCE EDUCATION CENTER
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS

PURPOSE

The improvement and maintenance of the quality of instruction secondary school inservice teachers receive is a continuing concern for science educators. One way of meeting this concern has been to help science teachers develop their teaching techniques by offering them several methods of inservice education including academic year and summer institutes, workshops, and self-administered materials. Of the several inservice methods mentioned, materials that are self-administered are unique in that they can be designed for use solely in the science classroom. The classroom is a special environment for inservice activity because students are there. With students present, teachers have the opportunity to implement different teaching techniques. And since students are frequently directly involved in classroom teaching practices such as discussion and laboratory activities, teachers also have a source of feedback about how they use specific classroom practices. The purpose of this study was to investigate the effectiveness of educational materials designed to be administered by science teachers in the classroom setting and used alone or with student feedback in modifying selected teaching practices.

RATIONALE

Field psychology provides a theoretical basis for suspecting that feedback from students about teaching practices and teacher self-assessment of teaching practices will affect teachers' behaviors. According to McCormack (1973), behavior is greatly influenced by the way people see themself in relation to their environment. Similarly, Combs (1965) has noted that peoples' behavior at any instant is the result of 1) how they see themselves, 2) how they see the situations in which they are involved, and 3) the interaction of these two.

In the classroom the "environment" for the teacher would in large part be the students. According to field psychology, teaching behaviors would be expected to be effected by both teacher self-assessments of behavior and by information from the environment related to the teacher or teaching, that is, by student behavior or students' perceptions about teaching practices (see Lewis, 1978, for a detailed discussion of the research in this area).

Gage, Runkel, & Chatterjee's (1963) results provide support for the theoretical position that teachers' knowledge of a discrepancy between what students perceive as the occurrence of behaviors of the "best imaginable teacher" and "achieved" behaviors promotes change in some of the teaching behaviors assessed. The change is towards the frequency of occurrence perceived for the "best imaginable teacher."

In studies by Bryan (1963), Tuckman & Oliver (1968) and Novak (1972), other types of feedback from students were utilized. Both Bryan (1963) and Tuckman & Oliver (1968) found feedback in the form of student ratings of how well a teacher performed various teaching practices or ratings of teacher characteristics to be effective in producing gains in ratings.

Novak (1972) investigated the effects of feedback in the form of students' reactions to instruction on the behaviors of eighteen biological science teachers. Audiotape analysis indicated that teachers receiving student feedback changed teaching behavior, generally, towards less teacher talk and lecture. A significant decrease in the mean numbers of questions asked each month was also found.

PROBLEM

From the studies above it can be concluded that some types of feedback from students may be effective in changing teaching behavior. Also, the feedback may be more effective in some situations and from some behaviors than for others. Beam and Horvat (1975) have pointed out that in general, teachers and students do not perceive the occurrence of certain teaching practices in the same way. Will modification in teaching behavior occur when teachers compare their perceptions with students' perceptions of the occurrence of certain classroom practices? Are teachers' self perceptions of the occurrence of teaching practices alone sufficient to change behavior? Will change in behavior occur when the assessment of teaching practices (including the collecting, scoring, and summarizing of the data) is carried out entirely by the teacher using written instructions and prepared assessment materials?

The study was designed to answer the following question: *Does teacher self-analysis with and without student feedback change the frequency of occurrence of selected teaching practices?*

RESEARCH DESIGN

Participants were secondary school science teachers who volunteered for the study. Data from 39 teachers participating in the study were utilized. Each of the participating teachers selected one class of science students to use in the study. No criteria for the class selection were given to the teachers. Over 69% (27) of the 39 classes were seventh and eighth grade science classes. The other classes included all levels from sixth grade through twelfth grade.

PROCEDURE

Teachers participating in the study were randomly assigned to treatment or control groups. Treatment Group I (N = 12) completed a self-analysis with pupil feedback. Treatment Group II (N = 15) completed a self-analysis and did not receive student feedback. Group III teachers (control) (N = 12) received no treatment.

Self-analysis included 1) completing a self-assessment of the frequency of occurrence of four teaching practices and 2) deciding to focus on two of the teaching practices for six weeks.

Teacher self-assessment of teaching practices is the frequency of occurrence of teaching practices, as perceived by the teachers, and determined by the completion of a questionnaire *(Science Classroom Practices Questionnaire — Teacher Form)* and the scoring and graphing of the questionnaire data.

Student feedback is the assessment of the frequency of occurrence of teaching practices, as perceived by students and determined by the completion of a questionnaire *(Science Classroom Practices Questionnaire — Student Form)* and the scoring and graphing of the questionnaire data.

Group I teachers' choices of behaviors on which to focus were made after both self-assessments and student assessments had been collected, scored and summarized on a histogram by the teachers. Group II teachers' choices of behaviors on which to focus were made after teachers collected, scored and summarized only the teacher self-assessment information.

All three groups collected student assessments at the start of the study (Initial Pupil Assessments) and at the end of the study (Final Pupil Assessments). Results of student assessments were to be seen only by teachers in Group I. Teachers in Group II and the control Group III were instructed not to look at students' Initial Assessments and returned these assessments to the researcher from one to three weeks after receiving the materials. Teachers in Group I and Group II completed Initial and Final Teachers Self-Assessments. Control Group III teachers did not complete any teacher self-assessments.

The Science Classroom Practices Questionnaire — Teacher Form and *The Science Classroom Practices Questionnaire — Student Form* were developed to assess the frequency of occurrence of four teaching practices. The four teaching practices assessed were TP_1) The teacher directs questions to many different students, TP_2) The teacher redirects students' answers or comments, TP_3) The teacher has students tell about things related to science they have brought to class, and TP_4) The teacher has students talk over their laboratory work with other students while they are doing an investigation. The four teaching practices were selected from a group of practices judged as "promoting student participation." TP_1 and TP_2 were judged as being verbal behaviors that promoted student participation; TP_3 and TP_4 were judged as promoting student participation through the teacher's allowing time for or planning an activity that involved students (referred to as "managerial" practices).

Change in teaching practices was determined from the data in the questionnaire completed by students at the start of the study (Initial Student Assessments) and again, six weeks later (Final Student Assessments). Change was inferred from the difference between the mean final and mean initial student assessments of teachers in treatment subgroups and the control group.

Teachers were subgrouped according to the teaching practices selected for focus. Table 1 shows the subgroups and the numbers of teachers and students in each subgroup. Subgroups are designated by a Roman numeral representing treatment or control group (I, II, III) and a subscript representing the teaching practice selected for focus (TP$_1$, TP$_2$, TP$_3$, TP$_4$). All teachers in treatment Group I who decided to focus on TP$_1$ were subgrouped together and designated subgroup I$_1$; those focusing on TP$_2$ were designated I$_2$ and so on. Since each teacher selected two teaching practices for focus, each teacher was thus a member of two subgroups. For example, Teacher X was assigned to treatment Group I and selected TP$_1$ and TP$_3$ for focus. Teacher X was therefore a member of subgroups I$_1$ and I$_3$.

Table 1

NUMBERS OF TEACHERS (CLASSES) AND STUDENTS

			No. of Teachers (Classes)		No. of Pupils Completing Assessment (Combined Classes)	
Treatment Group	**No. of Teachers (Classes)**	**Treatment Subgroup**	**Initial Assessment**	**Final Assessment**	**Initial Assessment**	**Final Assessment**
I	12	I$_1$	9	9	192	185
		I$_2$	3	3	57	59
		I$_3$	7	7	90	122
		I$_4$	5	5	109	105
II	15	II$_1$	5	5	128	123
		II$_2$	10	10	245	247
		II$_3$	6	6	67	77
		II$_4$	9	9	216	225
III	12	III$_1$	12	12	325	318
		III$_2$	12	12	325	316
		III$_3$	12	12	196	202
		III$_4$	12	12	321	314

Teachers in Group III did not receive a treatment and so did not decide to focus on certain teaching practices. Teachers in Group III are designated as III$_1$, III$_2$ and so on to indicate that performance of a particular teaching practice was being considered. Each of these groups (III$_1$, III$_2$, III$_3$, and III$_4$) included all teachers assigned to Group III (refer to Table 1).

The assumption was made that teachers intended to increase the frequency of occurrence of the practices selected for focus. Telephone interviews with the teachers supported this assumption. Thus, decreases in teaching practices from Initial to Final Pupil Assessments are not considered the intent of treatment group teachers. The treatments are to be considered "effective" if behavior is modified in the direction of increased frequency.

INSTRUMENTATION

The student and teacher questionnaire developed for the study each consisted of 27 items of which 17 items were statements or questions about the four teaching practices while 10 items were fillers (Lewis, 19). Each of the four teaching practices was measured by three to five items. All items were written in the form of a behavior observable to a student (or teacher, for the teacher version). Frequency of occurrence of teaching practice was determined using the format of a bipolar graphic scale. Examples of items for the Student's Questionnaire are given below:

13. How often does this teacher use the name of a student when he (she) calls on them to comment or answer a question during class discussion?
    always ______________________________ never
14. How often does this teacher ask students to say something about another student's answer to a question during class discussion?
    always ______________________________ never
15. How many times does this teacher ask students to agree or disagree with something another student has said during class discussion?
    5 times or more ______________________________ 0 times

To complete the questionnaire, the student places an "X" on the line indicating his perception of the frequency of the teaching practice. Item scores are obtained by assigning a number from one to nine to the location of the "X" marked by the individual. The number is determined by placing a transparent scoring stencil, having nine equal parts, over the line on the questionnaire.

Reliability of the student questionnaire was determined using data collected from 40 classes (39 teachers in Groups I, II and III and a nonparticipant). Each of 17 items on the questionnaire was related to one of the four teaching practices of concern and was in a subgroup with from three to five related items. The correlation between an item score and the score for its related subscale was determined using Veldman's Program TESTAT. Item-scale correlations for 17 items ranged from 0.62 to 0.88. For 10 of the 17 items, the range was from 0.70 to 0.78. Intraclass correlations provided information regarding judgmental consistency of the students. Correlation coefficients were determined using Veldman's Program INTRAR. As described by Veldman (1974):

> The intra-class correlation indicates the relative excess of among-subjects, as compared to among-raters variation, and therefore indicates judgmental consistency. The RI coefficient indicates the

> reliability of single judge ratings, while the RA coefficient estimates the reliability of the average ratings of a set of that many judges (p. 32).

Teachers in Groups I, II and III and a nonparticipating teacher were each assessed by one classroom of students. RI and RA coefficients were determined for initial student assessments and final student assessments. Results are summarized in Table 2.

Table 2

INTRA-CLASS CORRELATION COEFFICIENTS FOR INITIAL AND FINAL PUPIL ASSESSMENTS

	No. of Subjects	No. of Raters per Subject	Range RI	Av. RI	Range RA	Av. RA
Initial Pupil Assessment	40	35	0.05–0.34	0.06	0.57–0.92	0.74
Final Pupil Assessment	40	36	0.06–0.31	0.13	0.60–0.92	0.74

Results indicate that the average rating of teaching practices by a classroom of students is reliable, but not the rating of individual students. These results are consistent with those of Cogan (1958).

Instruction Booklets appropriate to each treatment and control group provided teachers with the procedure to follow during the six weeks of the study. Each Booklet contained an appropriate "Checklist" for teachers to follow during each week. Booklets also included, as appropriate, instructions for administering and scoring teacher and student assessments and pages to make calculations and graphs of assessment information.

Teachers in treatment Groups I and II received all appropriate materials in the first mailing; teachers in the control group (Group III) received only Initial Pupil Assessments questionnaires at first mailing; Final Student Assessment questionnaires were mailed to teachers in the control group during the fifth week of the study. Two mailings were necessary in an attempt to ensure that control group teachers did not have knowledge of the teaching practices being assessed by students.

Initial student assessment questionnaires for Groups II and III were returned to the researcher from one to three weeks following teachers' receipt of the materials. Teachers in Groups I and II were contacted the fifth week of the study to remind them that materials should be completed and returned. Teachers in Group III received their Final assessment forms the fifth week. Final Student

Assessment questionnaires of teaching practices were administered the sixth week of the study; all remaining materials were then returned to the researcher.

DATA ANALYSIS

To determine the effectiveness of the treatments in modifying the frequency of occurrence of the teaching practices, t ratios were calculated using Students Initial Assessments and Students Final Assessments. The t ratios provided a statistical test of the difference between gain or loss in mean student scores for treatment subgroups and the control group. The formula used to determine t was

$$t = \frac{\text{Difference of the difference of the means}}{\text{Standard error (SE) of the difference of the difference of the means}}$$

Two-tailed tests were applied to the t distribution since either positive or negative modification in the teaching practices was possible. Eight null hypotheses were tested; four of these concerned whether self-analysis with student feedback would modify each of the four teaching practices compared to no treatment; the other four hypotheses dealt with whether self-analysis alone would modify each teaching practice compared to no treatment. The null hypotheses were rejected for values of t equal to or greater than the value which would be obtained by chance 5% of the time. Treatments were considered "effective" if behavior was modified in the direction of increased frequency.

Table 3

COMPARISON OF GAINS BETWEEN TREATMENT AND CONTROL GROUPS FOR EACH TEACHING PRACTICE

Teaching Practice No.	Gains			Difference Between Gains	Standard Error of Difference Between Gains	t
	Self-Analysis/ Feedback Treatment	Self-Analysis Only Treatment	Control			
1	−0.232		−1.803	1.569	1.035	1.52
2	−3.387		1.263	−4.650	2.070	−2.25*
3	−0.158		1.146	−1.304	1.002	−1.30
4	2.457		0.247	2.432	1.180	2.06
1		0.021	−1.802	1.823	1.207	1.51
2		1.514	1.263	0.251	1.204	0.21
3		2.667	1.146	1.521	1.212	1.25
4		−0.662	0.247	−0.909	0.898	−1.01

*significant at .05 level (refer to Tables 4.1 - 4.12 for df's)

RESULTS

Table 3 compares the gains between treatment and control groups for each teaching practice.

Of eight null hypotheses concerning the effectiveness of self-analysis with and without student feedback, one was rejected. The results are summarized as follows:

1. Four hypotheses concerned the effectiveness of self-analysis with student feedback in modifying teaching practices as compared to the control group; one of the hypotheses was rejected. For teaching practice #2 (The teacher redirects students' answers and comments), the difference between initial and final assessments of the self-analysis/feedback subgroup was significant compared to that of the control group (refer to Table 4). The difference was in the negative direction. However, generalizations from this result had severe limitations since the score of a single teacher (out of three in the subgroup) was responsible for the difference. The score was larger than any difference score for any teacher participating in the study.

Table 4

DIFFERENCES IN GAINS OF THE SELF-ANALYSIS/FEEDBACK SUBGROUP AND THE CONTROL GROUP OF TEACHING PRACTICE #2

	Self-Analysis/ Feedback Treatment (n = 3)	Control (n = 12)	Difference	Standard Error	t
Mean (Final Assessment (N_T = 57) (N_C = 325)	22.525	26.494			
Mean (Initial Assessment) (N_T = 59) (N_C = 316)	25.912	25.231			
Gain/Loss (Final-Initial)	−3.387	1.263	−4.650	2.070	−2.25*

n = number of teachers (classes) participating in the initial and final assessments

N_T = combined number of pupils reporting for all classes in a treatment subgroup

N_C = combined number of pupils reporting for all classes in a control group

* significant at .05 level for df = 13 when t ≥ 2.16

No differences were significant between the gains of the feedback subgroups compared to the gains of the control group on assessments of teaching practices #1, #3, or #4. However, the difference between the gains on teaching practice #4 (The teacher lets students talk over their laboratory work while they are doing an investigation) was close to being significant in the positive direction (refer to Table 3).

2. Four hypotheses concerned the effectiveness of self-analysis alone in modifying teaching practices as compared to the control group; none were rejected. Differences in gains between teachers receiving self-analysis only and teachers in the control group were not significant for any of the four teaching practices.

CONCLUSIONS

The following is a summary of the conclusions of the study:

1. Self-analysis with student feedback and self-analysis alone were not effective in increasing the frequency of any of the four teaching practices assessed compared to the no treatment group.

2. There is some indication the self-analysis with student feedback may be more effective than no treatment in increasing the frequency that teachers let students talk over their laboratory work while they are doing an investigation.

Overall, then, one can conclude that the self-analysis process utilized in this study (where a teacher determined self-perceptions of the frequency of occurrence of four teaching practices and decided to focus on two practices) did not appreciably increase the occurrence of these same behaviors. Similarly, no increase in the frequency of the teaching practices occurred when teachers' self-analyses were compared with students' perceptions of the same teaching practices that teachers had collected and summarized. Increased occurrence was not apparent either for two "verbal" or two "managerial" teaching practices.

IMPLICATIONS

No differences between behavior of treatment and control group suggests that the comparison of teacher and student assessments of teaching practices as utilized in this study was not successful in producing a perceived change in teaching. The overall theoretical approach of self-assessment with student feedback should not, however, be abandoned as a viable means of changing behaviors. Alternative approaches such as allowing teachers more participation in the selection of the behaviors being assessed should be attempted. Also, the comparison of present behavior to idealized behavior may be the needed component for producing change that could be worked into a self-analysis procedure.

REFERENCES

Beam, K. J. and Horvat, R. E. Differences among teachers' and students' perceptions of science classroom behaviors, and actual classroom behaviors. *Science Education*, 59*(3)*; 333-334, 1975.

Bryan, R. C. Reactions to teachers by students, parents, and administrators. United States Office of Education, Cooperative Research Project No. 668. Kalamazoo: Western Michigan University, 1963.

Cogan, M. L. The relation of the behavior of teachers to the productive behavior of their pupils. *Journal of Experimental Education*, 27:89-124, 1958.

Combs, A. W. *The Professional Education of Teachers*. Boston, Mass.: Allyn and Bacon, Inc., 1965.

Gage, N. L.; Runkel, P. J.; and Chatterjee, B. B. Changing teacher behavior through feedback from pupils: An application of equilibrium theory. Readings in the *Social Psychology of Education.* (Edited by W. W. Charters, Jr., and N. L. Gage.) Boston, Mass.: Allyn and Bacon, 1963, pp. 173-181.

Lewis, M. C. *The effectiveness of pupil feedback and self-analysis in changing selected teaching practices of secondary school science teachers.* Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1978.

McCormack, A. J. Behavioral Objectives for science methods courses: A humanistic approach. *Science Education*, 57:55-64, January, 1973.

Novak, J. H. A study of the effects of the use of a pupil response instrument on the behaviors of biological science teachers. Final Report. Pittsburgh University, Pa., Pub. November, 1972. (ERIC ED 069 155, 1973).

Tuckman, B. W. and Oliver, W. F. Effectiveness of feedback to teachers as a function of source. *Journal of Educational Psychology*, 59*(4)*:297-301, 1968.

Veldman, D. J. *VSTAT user manual CDC 6600 computer programs for statistical analysis.* Austin, Texas: Ibid, Inc., 1974.

THE EFFECT OF RATING-DELAY, INSTRUCTIONAL CONTENT, AND LENGTH OF ATTENDANCE ON ELEMENTARY TEACHER RATINGS OF SCIENCE INSERVICE WORKSHOPS

THOMAS C. MEADOWS
DEPARTMENT OF CURRICULUM, MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF MANITOBA
WINNIPEG, MANITOBA, CANADA

INTRODUCTION

Inservice education has a history dating from the 1850's. It appears that continuing education needs of teachers are repeated in cycles, perhaps reflecting supply and demand or perhaps recurring needs for standards. Much of the literature surveyed has been descriptive in nature with few studies designed to test inservice variables, themselves. Questionnaire studies on teacher preference for inservice format were generally in agreement that the workshop and demonstration types of instruction be used. Both long and short periods of instruction were shown as capable of producing significant gains in teacher or student behavior, but were not of the type that deliberately manipulated a duration variable. One study (Spradlin, 1973) suggested that, for summer institutes, a follow-up of as long as two to three years might be important in obtaining the ratings of participants. But no study was identified that studied a rating-delay variable. The literature reviewed cites the importance and need to identify variables that contribute to successful inservice education designed for teachers (see Meadows, 1978).

PROBLEM

This study was designed to investigate the organization and structure of short, elementary science inservice workshops in an attempt to obtain answers to the following questions: Does the content of an inservice workshop affect a teacher's rating of that workshop? Does the length of an inservice workshop affect a teacher's rating of that workshop? Does the length of time between a workshop and the rating of the workshop affect a teacher's rating of that workshop? Are there interaction effects among the three variables of content, instructional length, and the length of time between a workshop and the rating of that workshop?

The independent variables were: 1) length of instruction; 2) content of instruction; and 3) the time interval between the workshop and the teachers' rating of the workshop (rating-delay). The dependent variable was the teachers' rating of the workshop based on its usefulness and on knowledge gained. This rating was based on the rating score obtained on a *Science Inservice Rating Scale (SIRS)* developed for the study (Meadows, 1978).

RESEARCH HYPOTHESES

As a result of careful analysis of the problem the following research hypotheses were generated:

Hypothesis 1. There is no difference in teachers' rating of an inservice workshop that can be attributed to the instructional content of the workshop.

Hypothesis 2. There is no difference in teachers' rating of an inservice workshop that can be attributed to the length of the workshop.

Hypothesis 3. There is no difference in teachers' rating of an inservice workshop that can be attributed to variations in the length of time after a workshop when the ratings are made.

Hypothesis 4. There is no interaction between the workshop content and the workshop length and the teachers' rating of an inservice workshop.

Hypothesis 5. There is no interaction between the workshop content and the variations in length of time when ratings are delayed and the teachers' rating of an inservice workshop.

Hypothesis 6. There is no interaction between workshop length and the variations in length of time of rating-delay and the teachers' rating of an inservice workshop.

RESEARCH DESIGN

A posttest only, 2x2x3 factorial design (Campbell & Stanley, 1963; Isaac & Michael, 1976) was employed in this study. There were two levels of an instructional *content* factor; two levels of a workshop *length* factor; and three levels of a *rating-delay* factor — the time-delay between a workshop and the rating of that workshop by teachers.

The two levels of the instructional content factor involved workshop instruction of 1) Curricula and Text-Overview; and 2) Program Guide Book Orientation. The two levels of the instructional length factor included: 1) a *one-day* length of workshop; and 2) a *half-day* length of workshop. The three levels of rating-delay between the workshop and its rating by teachers included: 1) *short* — a delay time of 3 to 5 weeks; 2) *medium* — a delay time of 6 to 14 weeks; and 3) *long* — a delay time of 15 to 30 weeks. This resulted in 12 experimental sub-groups for the study. In addition to the foregoing factors, the statistical treatment allowed the study of interactions between pairs of factors.

INSTRUMENTATION

The investigation was developed in several stages. These included piloting for the purpose of organizing and refining the instruction and content of both one-day and half-day workshops. At the same time the development and refinement of the rating instrument, *Science Inservice Rating Scale* (SIRS), was completed.

The variety of conditions for the study required an instrument capable of sensing diverse, and possibly slight, contributions to teachers' knowledge and classroom use made possible by the workshops. Since an instrument was not readily available, construction and testing of an instrument that reflected the goals and the directions of a changed Provincial curriculum was important.

The instrument was developed from a pool of 220 items which was finally reduced to 60 items by standard test development procedures (Meadows, 1978).

The resulting *Science Inservice Rating Scale* (SIRS) appeared to be both a valid and a reliable instrument for asking theoretical and practical elementary science inservice workshop questions. The instrument, as designed, could gather, efficiently, participant views of the instructional contributions obtained from a workshop by requiring teachers to estimate both their pre- and post-conditions. This latter technique produces a change estimate that is free from the absolute scale requirements necessary when two separately administered *before* and *after* or "gain score" tests are administered.

INSTRUCTOR TRAINING

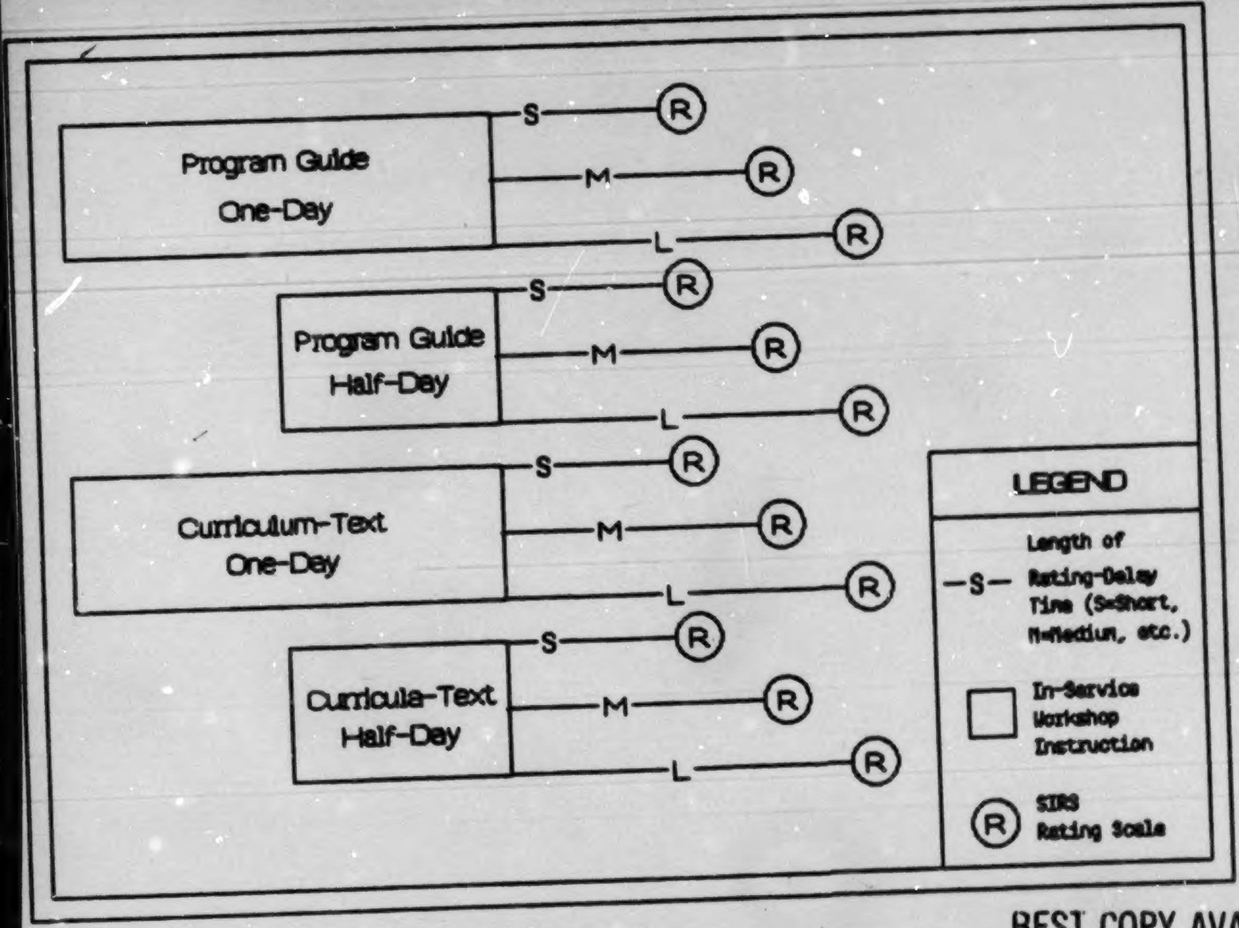
Instructors were trained for the teaching of either the half-day or the one-day workshops when they participated in a one-week comprehensive training session. Two such sessions were given, with 85 trainees attending either one of the two sessions. From the 85 trainees, 14 were selected for the study on the basis of suitable instructional uniformity and capability, and depending upon their being able to represent selected areas throughout the Province (Manitoba, Canada). The week of training provided trainees with a wide range of topics. These included: processes of science; student evaluation; use of the out-of-doors; materials and laboratory organization; subject integration; learning centers; philosophy and goals; resource leadership; and the use of the new program guide (being introduced throughout the Province along with a new elementary science curriculum).

Local conditions and needs determined the final content and length of workshops when these trainees returned to their respective divisions to offer the one-day and half-day workshops to their local teachers. A distinct difference in the content of two types of workshops was planned. The "curricula-test overview" was designed to be the more practical, with a great variety of practical and worthwhile ideas from the new text programs, with representative units illustrated in a workshop setting. The "program guide orientation" was designed to be more theoretical, philoso phical, and professional in content, including orientating of teachers to the use of the new program guide — its format, the "spirit of science" background, and other philosophical and nature of science positions as supported by the current learning theories of Piaget, Bruner, and others.

The two types of workshops were then given to meet the professional developmental needs of groups of teachers. Preliminary studies about these needs had shown that some area groups were more concerned about the curricula-text type of content (for text adoption purposes and the like), while other area groups were more concerned with the NEW program guide and its message.

The basic sequence and organization of the workshops and their differing lengths of rating-delay time are shown in Figure 1.

Figure 1
FLOW CHART SHOWING ORGANIZATION
AND SEQUENCE OF STUDY


Program Guide
One-Day
S
M
L
R
Program Guide
Half-Day
Curriculum-Text
One-Day
Curricula-Text
Half-Day
LEGEND
—S— Length of Rating-Delay Time (S=Short, M=Medium, etc.)
In-Service Workshop Instruction
R SIRS Rating Scale

BEST COPY AVAILABLE

Table 1

DISTRIBUTION OF TEACHERS IN TERMS OF PERSONAL AND CONTEXTUAL VARIABLES

Instruction			Locale			Instruction Level		Years of Experience			Gender		Educational Standing			Classroom Type		Preference of Teaching Subject			
Content and Rating Delay	Length in Days	Group Size (n)	Rural	Urban	Suburban	Primary	Intermediate	Long 11+	Medium 5-10	Short 1-4	M	F	High School Only	Under Grad	Degree or Beyond	Open Area	Self Contained	Science	Math	S.S.	L.A.
Curricula	One-Day																				
Long		18	9	—	9	9	9	1	9	8	5	13	1	10	7	5	13	2	8	1	7
Medium		16	5	—	11	5	11	5	8	3	7	9	3	6	7	8	8	3	4	4	5
Short		16	6	2	8	7	9	4	6	6	4	12	—	7	9	2	14	1	7	1	7
Curricula	Half-Day																				
Long		12	10	—	2	5	7	3	6	3	4	8	3	6	3	4	8	2	6	2	2
Medium		18	—	10	8	8	10	6	8	4	6	12	2	9	7	11	7	4	5	4	5
Short		19	—	16	3	8	11	7	6	6	4	15	1	7	11	9	10	2	7	3	7
Program Guide	One-Day																				
Long		13	6	—	7	8	5	3	4	6	4	9	1	5	7	5	8	4	—	2	7
Medium		18	3	2	13	9	9	6	9	3	6	12	4	4	10	8	10	—	7	5	6
Short		19	2	8	9	8	11	6	6	7	7	12	—	8	11	13	6	6	4	3	6
Program Guide	Half-Day																				
Long		11	10	—	1	3	8	3	5	3	5	6	1	6	4	1	10	3	2	2	4
Medium		12	4	2	6	5	7	1	5	6	3	9	2	3	7	7	5	3	5	1	3
Short		16	2	10	4	7	9	5	5	6	4	12	2	8	6	13	3	3	4	3	6
TOTALS-EXPERIMENTAL:		188	57	50	81	82	106	50	77	61	59	129	29	79	89	86	102	33	59	31	65

BEST COPY AVAILABLE

SAMPLE

The teachers participating in the study were from schools selected from rural, suburban, and urban areas of Manitoba. The selection of instructors was made so as to include a representative cross-section of teachers typical of the province as a whole. A stratification procedure was used to have participants from at least three instructors in each cell group. This was simplified, to some extent, by assigning participants of each instructor to the three different levels of rating-delay. The procedure tended to minimize the effects of both the demographic and the instructor variables. Finally, a decision whether to include any one of the centers taught by any one instructor made it possible to provide a degree of "randomization" in that any of the centers could equally have been chosen. Only those teachers willing to co-operate, however, were participants in the final return mailings.

The result was 188 subjects distributed in the various experimental groups as shown in Table 1. There were 57 rural, 50 urban, and 81 suburban teachers divided somewhat equally between primary and intermediate levels, with twice as many female as male teachers. Slightly more than half the sample taught in self-contained classrooms. Almost equal numbers of the participants preferred science and mathematics teaching as compared to those who preferred social studies and language arts. The teachers were almost equally divided in terms of teaching experience. The largest cell in the study contained 19 teachers, and the smallest cell 11 teachers with an average cell distribution of 15.8 teachers. This was judged to be suitably representative of the teaching population of Manitoba's schools.

A further check on bias in the sample was made in terms of the attrition of subjects that occurred between the mailings to the schools and the final returns of the SIRS booklets. Percentage returns calculated for each of the cells used in the investigation indicated a mean response of 60.8 percent with attrition in the levels of the three main factors varying by no more than 4.0 percent. This also suggested that there were no systematic differences among the levels used to test the various hypotheses.

As previously discussed, the basic sequence and organization of the workshops and their differing lengths of rating-delay time are shown in Figure 1. The variation or spread in the dates or workshops, while not shown in the flow chart, occurred during four months of the spring term. Similarly, the number of inservice workshops were offered throughout the four months. Subjects attended only one of these workshops, and rated the workshop at only one of the three possible times. The trainee instructors, of course, offered one or the other of the workshops as many times as the occasion demanded and frequently, in more than one division or school area. The returned rating scales were grouped and analyzed in terms of their respective workshop types (one-day or half-day lengths; program guide or curriculum text contents, and short, medium, or long rating-delay time periods).

DATA ANALYSIS

Data collected from the study were analyzed first using analysis of variance (ANOVA) on a computer program designed to analyze multi-factorial designs with equal or unequal sub-class numbers. A posteriori means-contrasts were done using the Scheffe method (Winer, 1971). All hypothesis testing was done at the 0.05 level of significance. The Cochran C Test (Roscoe, 1975) was used to test the homogeneity of variance for the ANOVA. Correlations were calculated with a computer program using a Multiple Linear Regression program with ANOVA.

RESULTS

A three-factor ANOVA of the scores, plus comparisons among the means using the Scheffe test, were the bases for the interpretation of the data. The findings are organized in terms of the six main hypotheses stated for the study. The complete ANOVA is summarized in Table 2.

Table 2

ANALYSIS OF VARIANCE SUMMARY TABLE FOR ALL FACTORS AND FOR FACTOR INTERACTIONS

Source of Variance	df	Sum of Squares*	Mean Square	F Ratio	Probability
Length	1	2845.6	2845.6	16.33	$p < .001$
Content	1	2551.8	2551.8	14.64	$p < .001$
Rating-Delay	2	2360.9	1180.5	6.77	$p < .01$
Length x Content	1	361.3	361.3	2.07	n.s.
Length x Delay	2	1732.7	866.4	4.97	$p < .05$
Content x Delay	2	730.1	365.1	2.09	n.s.
L x C x D	2	821.4	410.7	2.36	n.s.
WITHIN	176	30672.2	174.27		
TOTAL	187	43110.7			

*Sum of Squares adjusted for unequal subclass numbers by the use of Harmonic Means of N (Snedecor).

A significant F-ratio was obtained for the three main factors of the ANOVA. There was also a significant interaction effect between Length of Instruction and Rating-Delay. Subsequent tables provide the results of means testing without the repeating of the F-ratio information given in Table 3.

1. *Content of Instruction*. For Hypothesis 1, the analysis of the data indicated a significant instructional content effect. The first research hypothesis was rejected. The data supported the position that the content of an inservice workshop affects the teachers' ratings of that workshop. The results from the comparison between the means of the curriculum-text overview content versus the program guide orientation content are given in Table 3. The mean score of 37.00 for the curricula-text overview exceeded the mean of 29.62 for the program guide workshops by a difference of 7.38. This difference was significant. The curricula-text overview provided higher teacher ratings than did the program guide orientation workshop content.

Table 3

SCHEFFE TESTS AMONG THE MEAN SCORES OF THE CONTENT OF INSTRUCTION FACTOR

Comparision	n*	Mean Score	Difference	k/df	Scheffe F	Probability
Curricula-Text	99	37.00				
Program Guide	89	29.62	7.38	2/1,176	14.65	p < .001

*Harmonic Mean = 93.73

2. *Length of Instruction*. For Hypothesis 2, the analysis of the data indicated a significant instructional length effect. The results of the means-comparisons between the one-day and the half-day instructional lengths are given in Table 4. The one-day instructional length produced a mean score of 37.21, as compared to a mean of 29.41 for the half-day length. The difference of 7.79 was significant. The one-day workshop SIRS ratings by teachers were significantly higher than the half-day scores. Therefore the hypothesis was rejected.

Table 4

SCHEFFE TESTS AMONG THE MEAN SCORES OF THE LENGTH OF INSTRUCTION FACTOR

Comparision	n*	Mean Score	Difference	k/df	Scheffe F	Probability
One-Day Length	100	37.21				
Half-Day Length	88	29.41	7.79	2/1,176	16.31	p < .001

*Harmonic Mean = 93.67

3. *Length of Rating-Delay.* For Hypothesis 3, related to the *short* (3 to 5 weeks), *medium,* (6 to 14 weeks), and *long* (15 to 30 weeks) periods of delay in obtaining teachers' ratings, the analysis of the data indicated a significant effect for the lengths of rating-delay. These data supported the rejection of the third hypothesis. The length of time between a workshop and the rating of the workshop does affect teachers' ratings of the workshop. The results for the comparison among the means for *short, medium,* and *long,* lengths of delay are given in Table 5.

Table 5

SCHEFFE TESTS AMONG THE MEAN SCORES OF THE FACTOR OF LENGTH OF RATING-DELAY

Comparision	n*	Mean Score	Difference	k/df	Scheffe F	Probability
Long Delay	54	35.29				
and						
Medium Delay	64	36.34	1.05	3/2,176	.09	n.s.
Long Delay	54	35.29				
and						
Short Delay	70	28.31	6.98	3/2,176	4.26	p < .02
Medium Delay	64	36.34				
and						
Short Delay	70	28.31	8.03	3/2,176	6.19	p < .005

*Harmonic Mean of each pair used for comparison.

These results indicate that teachers' rating scores for an inservice workshop are not as great immediately following the workshop as are the scores after several weeks time-lapse. A graphic representation of the delay effect is shown in Figure 2. The mean scores are shown for each of the three delay levels. The higher mean scores were obtained by the medium and longer delay groups. The value of the ratings of the inservice workshops increases with time and then appears to decrease, somewhat. A rating after 5 weeks apparently allows for more favorable responses from inservice recipients.

4. *Interaction Effects.* There were no significant interaction effects between instructional content and length of instruction and, similarly, there were none between the instructional content and rating-delay. Therefore hypotheses 4 and 5 were rejected. Between the *Length of Instruction* and *rating-delay* there was a significant interaction effect. Therefore, the sixth hypothesis was rejected.

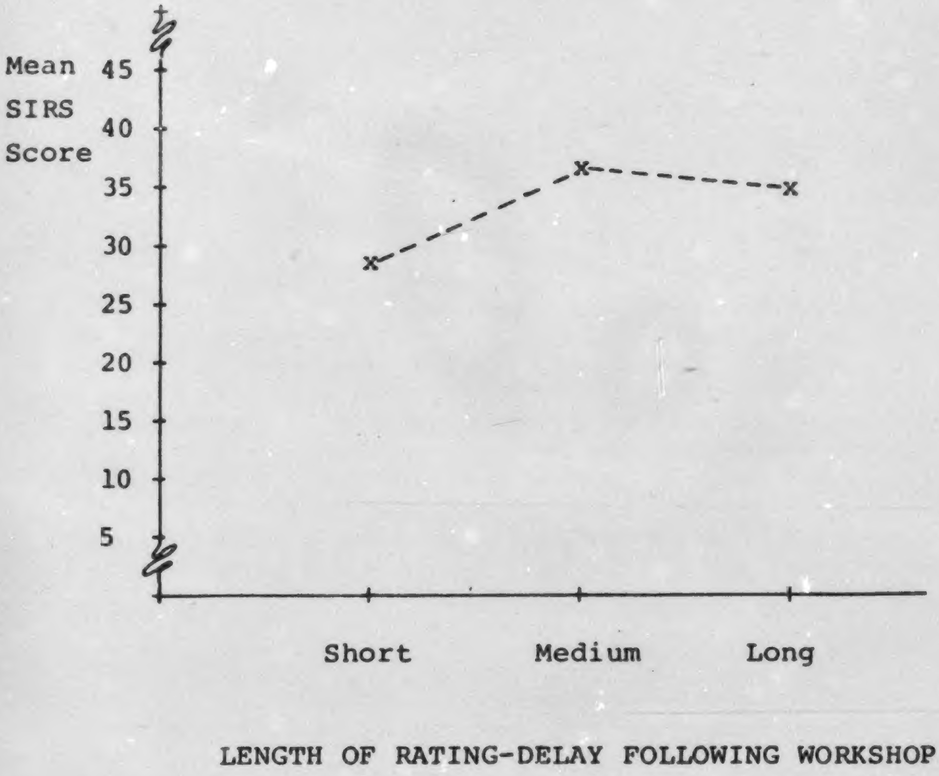
Scheffe tests among the several appropriate means of the one-day and the half-day instructional lengths in combination with the long, medium, and short rating-delay times, showed only one significant interaction effect. This was between the *Half-day* length of instruction and the *Long* rating-delay time. This was interpreted to indicate significant interaction between the *half-day* and the *long* delay levels for the two factors of length of instruction and rating-delay. This finding would support decisions not to use long-delay rating for half-day workshops. This was further supported by rating sheets returned with the comment that "they had forgotten," and the like. Undoubtedly, the longer workshop made for less forgetting.

DISCUSSION

The findings resulting from the analysis of data were used to answer the four study questions.

Figure 2

GRAPH OF THE EFFECT OF RATING-DELAY UPON THE TEACHERS' RATING OF WORKSHOPS.



QUESTION ONE

Does the content of an inservice workshop affect a teacher's rating of that workshop?

Content-of-instruction information given in the literature predicted, in part, some of the answers to this question. Educators argued for more philosophical and professional agendas to enable teachers to understand the rationale for the curricula they are required to teach, while reported research studies, on the other hand, indicated that teachers wanted to attend the more practical teaching-skill type of program. In this study, evidence was found to support that teachers rate the knowledge and classroom use of curricula-text/instructional content workshops significantly higher than the program guide/teaching rationale type of workshop content.

QUESTION TWO

Does the length of an inservice workshop affect a teacher's rating of that workshop?

Based on the SIRS ratings by teachers there was evidence to support that the length of an inservice workshop does affect the teacher's rating of the workshop, with the longer, one-day workshops producing greater teacher knowledge and classroom application than the shorter half-day workshops. The findings related to length-of-workshops require further interpretation because of the interaction effect between the long day levels of the rating-delay factor and the one-day and half-day length of workshop. This is discussed under questions three and four.

QUESTION THREE

Does the length of time between a workshop and the rating of the workshop affect a teacher's rating of that workshop?

Evidence was found to support that the medium or the longer period of rating delay (as defined in this study) is the preferred time for rating inservice workshops as designed for this study. The later evaluations produce greater SIRS scores — that is, teachers rate them as producing greater knowledge and classroom use for the teacher. The significant interaction occurring between the instruction length and the medium and long rating-delay periods requires that generalizations about the later evaluation periods should be made with great caution. There was evidence that the shorter delay period may be more appropriate with shorter instruction periods, but supports the use of the longer delay for the one-day workshops.

QUESTION FOUR

Are there interaction effects among the three variables of content, instructional length, and the length of time between a workshop and the ratings of that workshop?

With respect to length and content, there was no evidence to support that, for the instructional lengths and the two types of content, either instructional *length* was more preferable than the other for either of the two types of instructional *content;* nor was there evidence to support that the ratings of one content would be more affected than the ratings of the other content by one or other of the three delay periods. The effects of the significant interaction of length x rating-delay have been discussed in relation to each of the appropriate main effects and are not discussed here.

CONCLUSIONS

Conclusions are discussed for each of the major study questions. The SIRS instrument, developed for the study, is also discussed.

WORKSHOP CONTENT

It can be concluded that elementary school teachers rate the practical curricula-text-overview workshop content as giving more knowledge and classroom application than does a content of purpose, planning, and rationales. It was also concluded that different workshop contents will affect teachers' ratings of inservice workshops in terms of the teachers' knowledge and classroom application.

WORKSHOP LENGTH

Elementary school teachers ratings of the knowledge and classroom application obtained from an inservice workshop are different for differing lengths of workshop as rated on the SIRS instrument.

LENGTH OF RATING-DELAY

Longer rating-delay times produced higher mean scores than did the short delay period. While teachers did perceive knowledge change and classroom application shortly after a workshop, a further period of time appeared to enhance the ratings. The length of rating-delay interacted with the length of workshop such that the longer workshop produced significantly higher scores after the longer delay period, with the shorter workshop being significantly more favorably rated if the shorter delay period was used.

THE SIRS INSTRUMENT

It can be concluded that:

1. The SIRS instrument is capable of distinguishing between a workshop of a half-day length as compared to a one-day length of workshop.

2. the SIRS instrument also is able to show differences between two types of workshop content.

3. an instrument capable of distinguishing between very short elementary science inservice workshops has been developed.

4. the SIRS instrument is administratively useful. The instrument requires only one administration time. Like an interview, it asks teachers to consider items in association with the workshop in question. Teachers are not interrupted in their classrooms, but can complete the booklet at a convenient time, at one sitting, without interference or distraction.

5. the SIRS instrument can be used even after other workshops have intervened, for example, a language arts session. (This property of the instrument was discovered during the pilot studies when such a situation was encountered). Teachers are able to associate successfully the source of their science knowledge and classroom practices. In this regard, the SIRS instrument is much like a standardized interview form with the teacher being asked to associate in relation to each item answered.

6. the SIRS instrument has a very high reliability and has good validity of both content and concurrent types.

IMPLICATIONS

From teachers' points of view, the short inservice workshop does make changes in knowledge and classroom practices. Arguments that the short inservice workshop is not effective can be countered with the evidence as gathered from many teachers. Even teachers who indicated a dislike of some portion of a workshop were able to show knowledge changes or classroom applications gained.

The study of inservice workshops, in terms of slight differences in length, in terms of minor differences in content, and in terms of teachers evaluating by means of classroom feedback on the knowledge and use of inservice content, will require methods of detecting these differences. The method and direction established with the SIRS instrument may provide the means to make such study possible.

Because the court of instruction does affect teachers' knowledge and classroom use, there is a need for further research into what topics will produce optimum teacher change. Similarly, there is a need to discover the shortest workshop length that can produce, efficiently, an optimum change.

Further research is needed to determine, to a much more detailed extent, the optimum times of rating-delay for differing contents and lengths of instruction. The rather broad scale of delay, used operationally for this study and the interaction with length of workshop as encountered, limit possible generalizations at this time. Spradlin's (1973) study for summer institutes suggested follow-up that might be as long as two or three years! This aspect will need careful evaluation in terms of short inservice workshops, since teachers polled at the very long periods of delay time for this study did not react favorably to being asked to specifically remember what had transpired at a short workshop. The effectiveness of other review techniques, as compared to the SIRS method might, of course, remove this reaction and should be the focus of further study.

REFERENCES

Meadows, T. D. *The Effect of Selected Variables on Elementary Teachers' Ratings of Science Inservice Workshops*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1978.

Roscoe, J. T. *Fundamental Research Statistics for the Behavioral Sciences* (2nd ed.). New York: Holt, Rinehart and Winston, Inc., 1975, 290.

Spradlin, S. D. *A Study of Change in Science Teaching Behavior in the Classroom Following Involvement in Institute Training or Retraining Programs*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, July 1973.

Winer, B. J. *Statistical Principles in Experimental Design* (2nd ed.). New York: McGraw-Hill Book Co., 1971, 199, 570.

THE EFFECT OF A PERSUASIVE COMMUNICATION ON THE ATTITUDE TOWARD ENERGY USE OF PRESERVICE ELEMENTARY TEACHERS

THOMAS R. KOBALLA
SCIENCE EDUCATION CENTER
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS

INTRODUCTION

Attitude toward specific fuels and attitude toward energy use can be viewed as mutally exclusive dimensions of one's attitude toward energy (La Follette, 1980). Attitude toward specific fuels concerns such issues as, "Is damming the local river the best way to generate electricity?," "Should the regional nuclear power plant continue to be funded?," and "What is the best way to control coastal oil spills?" Viewed as peripheral in the daily lives of most Americans, the concerns addressed by this dimension of attitude toward energy are normally left to be resolved by governmental bodies. Conversely, concerns related to one's attitude toward energy use are not relegated but maintained. Thoughts and actions regarding the use or misuse of energy resources are affected by the attitude one holds toward energy use (see Ajzen & Fishbein, 1980).

PROBLEM

Realizing the importance of one's attitude toward energy use can influence personal thoughts and actions as well as the thoughts and actions of those readily influenced by that person (Lewin, 1951; Zimbardo, Ebbesen, & Maslash, 1977). School administrators and government officials can no longer allow chance to be the factor that determines the energy use practices of America's future generations. Treatments must be devised and tested that can favorably influence the attitude of teachers toward energy use.

A number of treatments have been devised to influence the attitude of teachers toward energy use (Barrow, 1982; Glass; 1982; Lawrenz, 1982). The length of these treatments ranges from weeks to months with mixed results. Involving teachers in such lengthy treatments seems ludicrous if in fact attitude toward energy use can be changed in minutes and that change resists dissipation.

RESEARCH QUESTIONS

Therefore, this investigation was designed to test the effect of a thirteen minute persuasive communication, systematically prepared to change attitude toward energy use of elementary teachers. Two questions identify the parameters of this investigation:

1. Can attitude toward energy use be positively changed as a result of exposure to a persuasive communication developed following Hovland, Janis, and Kelley's four part model?

2. Do positive attitude shifts between pre- and posttest, if any, resist dissipation three weeks after exposure to a persuasive communication?

RATIONALE

In a series of successive articles, Shrigley (1978, 1980, 1983) outlined how Hovland, Janis, and Kelley's (1953) persuasive communication model can be used as a vehicle for attitude change in science education. He pointed out how more than just sharing up-to-date information with teachers is required for the success of this stimulus-response approach. Successful attitude change, as Shrigley suggests, requires the acceptance of the opinions or conclusions of the message.

Since stimulus-response learning theory, stock-in-trade for educators, and the persuasive communication model, outlined by Shrigley, share theoretical undergirding, persuasive communication seems to be a logical avenue for studying attitude change in an educational context. Central to Hovland, Janis, and Kelley's model is the confrontation of subjects with knowledge suggesting a change in attitude. And initially this appears little different from conventional classroom instruction where one's goal is to get the subject's attention and endeavor to assure his comprehension. However, as interpreted for this investigation, the persuasive communication contained specialized information, a message laden with beliefs, in this case, beliefs about energy use. As suggested by Zimbardo, et al. (1977), it is this belief component of the message that is responsible for the subject's acceptance, or yielding to the information, that can bring about attitude change.

Hovland, Janis, and Kelley's four part summary question, "*Who* says *what* to *whom* with *what effect?*," provided the theoretical undergirding for this investigation. Within the context of the investigation the essential components of the summary question were interpreted as follows:

1. "Who" is the credible science educator instructing preservice elementary teachers in the importance of energy use.
2. "What" is the content of the persuasive communication, the belief suggesting the topic of energy use be an important part of the elementary school curriculum.
3. "Whom" is the recipient of the communication, preservice elementary teachers.
4. "What Effect" is the resultant attitude change.

RESEARCH DESIGN

The "Who" and "What" components of Hovland, Janis, and Kelley's four-part summary question were considered the stimulus of the S-R persuasive communication model. Two responses were collected from each of 113 preservice elementary teachers who responded to a qustionnaire asking them to state what credible characteristics they would associate with a science educator teaching energy use practices in a college science methods course. Considered peripheral, thus not incorporated into the persuasive communication, were categories that received a response rate equal to or less than 5 percent. The characteristics reaching criterion can be seen in Column A of Figure 1. The "What" component of the communication was identified from the responses of 36 science educators who shared two statements that they had used or would use in college science methods courses to persuade preservice elementary teachers of the need to teach energy use practices to children. The 72 believe statements collected were assembled into 12 categories by a jury of four science educators. The seven categories having equal to or greater than a 5 percent response rate, thus incorporated into the persuasive communication, can be seen in Column B of Figure 1.

INSTRUMENTATION

The measurement instrument represented Hovland, Janis, and Kelley's fourth and final component. The 22 item Energy Conservation Attitude Scale (Koballa, 1981) was used to measure the attitude toward energy use of the sample. That a majority of Americans view energy conservation as the least expensive and most easily achieved means of grappling with concerns of future energy availability and use justified the instrument's use (Rayborn & Tuten-Pucket, 1980; Simonis, 1982). Over the three administrations, scores on the attitude scale ranged from 43 to 94. Coefficient alpha's for the scale were reported at .85, .87, and .86 for the three administrations.

PROCEDURE

One hundred thirty-one students enrolled in several sections of an elementary science methods course at a large eastern university during the spring semester, 1981, participated as subjects. A pretest, posttest, delayed-posttest, control group design was employed with subjects randomly assigned to either of two groups. The sample represented Hovland, Janis, and Kelley's "Whom" component.

Following the random assignment of subjects, the Energy Conservation Attitude Scale was administered for the purpose of assessing pretreatment attitude. Three weeks later, subjects were exposed to either the persuasive communication treatment (T1) or the placebo treatment (T2). The persuasive communication treatment consisted of a videotaped persuasive communication, the core of which was the four characteristics of a credible communicator, as identified by the 113 preservice elementary teachers, and the seven belief statements, categorized from the responses of the 36 science educators. The duration of the persuasive communication treatment was 13:25 minutes. The placebo treatment was a 13:10 minute segment of the videotape "Investigating Siphons," a lesson from *Investigative Science in Elementary Education* (Pennsylvania Department of Education, 1975), that is no way eluded to the perceived importance of wise energy use.

Absenteeism during the course of the investigation resulted in incomplete data for 11 subjects. As a result these data were eliminated from subsequent analysis.

RESULTS

Noting a substantial difference between group means on the pretest, difference measures were employed in order to reduce the inequalities among groups not caused by treatment (Roy, 1966). See Table 1.

Table 1

MEANS AND STANDARD DEVIATIONS FOR PERSUASIVE COMMUNICATION TREATMENT AND PLACEBO TREATMENT AT EACH ASSESSMENT

Treatment	Pretest $\bar{X}$ / SD	Posttest $\bar{X}$ / SD	Delayed Posttest $\bar{X}$ / SD
Persuasive Communication (T1)	79.03 / 4.03	82.41 / 4.39	81.06 / 5.81
Placebo (T2)	74.23 / 4.46	72.92 / 5.21	72.77 / 6.21

Figure 2

GRAPHIC DEPICTION OF GROUP MEANS FOR PERSUASIVE COMMUNICATION (T1) AND PLACEBO (T2)

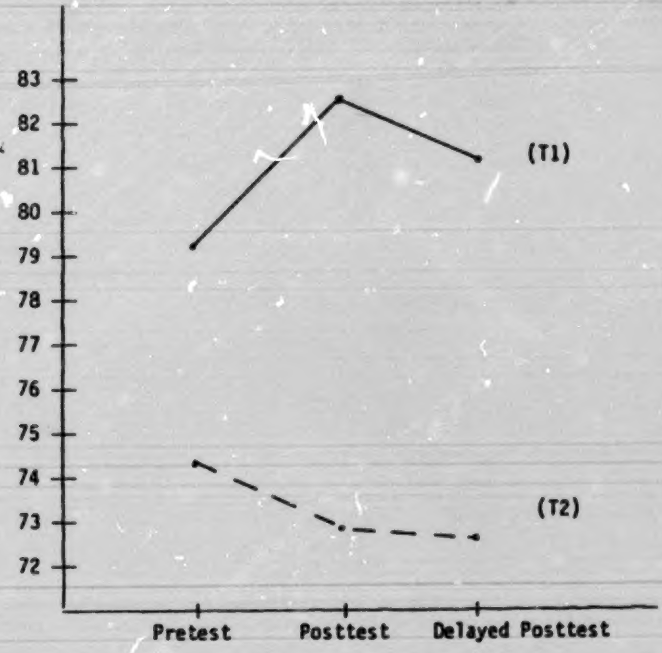


Figure 1

HOVLAND'S PERSUASIVE COMMUNICATION MODEL

Stimulus (S)			**Response (R)**
(A) WHO	**(B) says WHAT**	**(C) to WHOM**	**(D) with WHAT EFFECT**
A science educator is credible if he/she: 1) Is knowledgeable about energy conservation. 2) Demonstrates personal enthusiasm for energy conservation. 3) Demonstrates conservation principles by personal actions. 4) Adequately discusses conservation practices applicable to the lives of young children.	The belief statements of a persuasive communication on energy conservation provided by survey of science educators: 1) Most of our energy resources are non-renewable; the are finite. 2) Conservation of energy buys time for alternatives to be designed. 3) Conserving energy is a morally and economically sound practice. 4) Excesses in energy consumption will have a negative effect on the lives of future generations. 5) Generating accurate information is necessary for wise decision-making as well as basic to teaching conservation to our youth. 6) The conservation of energy is integrally related to the total scope of global environmental concerns. 7) The conservation of energy needs to be universal habit developed in mankind.	Preservice elementary teachers.	The change in teacher attitudes as measured by the Energy Conservation Attitude Scale.

To determine if attitude toward energy use can be changed using a brief, persuasive communication, a *t*-test for independent samples was calculated on the criterion of attitude change. Comparison between the mean difference cores (posttest minus pretest) of the persuasive communication (T1) and the placebo (T2) revealed a significant difference, $t(118) = 4.54$, $p \leq .01$ (see Table 2).

Table 2
COMPARISON OF TREATMENTS SUMMARY TABLE

Comparison	T1 $\bar{D}$ / SD	T2 $\bar{D}$ / $\overline{SD}$	df	t (two tailed)
Posttest - Pretest	3.38 / 6.54	−1.31 / 4.59	118	4.54*

* significant at the .01 level

To determine whether the positive attitude shift dissipated, a *t*-test for correlated samples was calculated. Comparison between the posttest and delayed posttest scores for the persuasive communication (T1) revealed a significant difference, $t(59) = -2.13$, $p \leq .05$ (see Table 3).

A posteriori comparison was carried out between pretest and delayed posttest means for the persuasive communication (T1) to see if there remained a significant difference between delayed posttest and pretest mean scores. The comparison revealed a significant difference between the group means, $t(59) = 2.03$, $p \leq .01$ (see Table 3).

Table 3
ATTITUDE DISSIPATION SUMMARY TABLE FOR SYSTEMATICALLY DESIGNED PERSUASIVE COMMUNICATION

Comparisons	$\bar{D}$ / $\overline{SD}$	df	t (two tailed)
Delayed Posttest - Posttest	−1.35 / 6.07	59	−2.14*
Delayed Posttest - Pretest	2.03 / 5.68	59	2.74**

*significant at the .05 level
**significant at the .01 level

CONCLUSIONS

1. A brief, systematically designed persuasive communication can favorably alter the attitude of preservice elementary teachers toward energy use.
2. The mean attitude score of the group while dissipating somewhat, remained significantly different from the group's pretreatment attitude three weeks following exposure to the persuasive communication.

3. Hovland, Janis, and Kelley's model, as operationalized in this investigation, seems to be a viable process for changing attitude in a schooling context.

DISCUSSION

While the purpose of this investigation was not to compare the developed persuasive communication with other treatments purported to change attitude toward energy use, it illustrated that the use of a systematically designed persuasive communication can positively alter the attitude of preservice elementary teachers. In addition, this investigation, testing attitude toward energy use of a sample of preservice elementary teachers, and that of Grabowski, et al. (1980), testing attitude toward science using a sample drawn from the sample population, resulted in similar findings. Hovland, Janis, and Kelley's four-part question, "Who says what to whom with what effect?" was the rationale over which both investigations were superimposed. Both were similar persuasive communications with credibility of the communicator and the belief statements of the message systematically designed as implied by Hovland, Janis and Kelley's model.

The consistency of the results of the two studies suggested that the probability of the reported attitude changes were not the result of change. The consistency also suggests that researchers need to begin to examine Hovland, Janis, and Kelley's model, as operationalized in these investigations, for testing other attitude objects important within the population of preservice, and perhaps inservice, elementary teachers.

The brevity of the treatment period of both studies could be an initial thrust in refuting a common, but until recently untested, bias among educators that attitude change required weeks or even months of treatment.

Critics of the Grabowski, et al. investigation, which had no delayed posttest, suggested that the treatment effect would dissipate shortly after the treatment. This investigation, however, employed a delayed posttest to check for evidence of dissipation three weeks after exposure to the persuasive communication. While the mean score of the group dissipated somewhat within three weeks following exposure to the persuasive communication, it remained significantly higher than the pretest mean. This finding suggests that mean positive shifts in attitude resulting from exposure to persuasive communications developed using Hovland, Janis, and Kelley's model tend not to dissipate significantly, at least not within weeks following the treatment. Of course, compatible results from investigations with the same and longer periods between posttest and delayed posttest are necessary before the dissipation factor can be ignored.

REFERENCES

Ajzen, I., & Fishbein, M. *Understanding attitudes and predicting social behavior.* Englewood, NJ: Prentice-Hall, 1980.

Barrow, L. H. Energy knowledge and attitudes and locus of control of secondary teachers. Paper presented at the National Association for Research in Science Teaching Annual Meeting, Fontana, WS, 1982.

Glass, L. W. An inservice energy education program for elementary school teachers. *Journal of Research in Science Teaching*, 1982, *19*, 469-474.

Grabowski, F. L., Shrigley, R. L., & Welliner, P. B. The systematic design of a persuasive communication for changing attitudes toward science. In M. R. Sunonsori and D. Rohner (Eds.), *Procedures of Selected Research Paper Presentations.* Washington, D.C. Association for Educational Communication and Technology, 1980.

Hovland, C. I., Janis, I. L., & Kelley, H. H. *Communication and persuasion.* New Haven: Yale University Press, 1953.

Koballa, T. R., Jr. *Designing a Likert-type scale to assess attitude toward energy conservation.* Paper presented at the annual meeting of The National Association for Research in Science Teaching, New York, April 1981.

LaFollette, M. C. Understanding, ignorance, and euphemisms: Reflections of the 146th annual meeting of the American Association for the Advancement of Science, January 1980. *Science Technology & Human Values*, 1980, *36*, 23-31.

Lawrenz, F. P. *Impact assessment of the 1981 portal school program.* Tempe, AZ: Evaluation Report Portal School Program-ASV, 1982.

Lewin, K. *Field theory in social science.* New York: Harper, 195.

Rayborn, S., & Tuten-Pucket, K. Use the newspaper to teach energy. *Science and Children*, 1980, *18*, 12-14.

Roy, W. S. *An introduction to experimental design.* New York: MacMillan Company, 1966.

Shrigley, R. L. The attitude concept and science teaching. *Science Education*, 1983, 67(4): 425-442.

Shrigley, R. L. The persuasive communication model: A theoretical approach for attitude change in science education. *Journal of Research in Science Teaching*, 1978, *15*, 335-341.

Shrigley, R. L. Science supervisor characteristics that influence their credibility with elementary school teachers. *Journal of Research in Science Teaching*, 1980, *17*, 161-166.

Simonis, D. A. Energy: The analogy approach. *The Science Teacher*, 1982, *49*, 41-44.

Zimbardo, P., Ebbesen, E. B., & Maslash, C. Influencing attitudes and changing behavior. Reading, MA: Addison-Wesley Publishing Company, 1977.

A COMPARISON OF METHODS FOR MEASURING THE DEGREE OF IMPLEMENTATION OF AN INNOVATIVE SCIENCE PROGRAM

JAMES D. ELLIS
BIOLOGICAL SCIENCES CURRICULUM STUDY
THE COLORADO COLLEGE
COLORADO SPRINGS, COLORADO

INTRODUCTION

The decade of the 1970's has been a period of disillusionment for science education. Yager (1980) heralds "a crisis in science education." Rowe (1980) describes the alarming status of elementary science as a vanishing species. Joseph Schwab indicated to Lee Shulman in an interview that the new science curriculum failed to change the traditional methods of teaching to that of an inquiry approach, but did succeed in updating the subject matter (Tamir, Blun, Hofstain, & Sabar, 1979). However, there is much to be learned from our successes and failures in science education. The future status of science education in the American society depends in part on substantiating the importance of our product and the effectiveness of our process.

Research in education during the late 1970's included a new focus on the implementation phase of curriculum change. This focus in implementation was in response to an increased awareness of failures to achieve the anticipated outcomes, desired educational benefits, and maximum implementation in large scale curricula projects (Gross & Herriott, 1979). Surely, if the public is to be expected to support major science education curriculum development efforts, the program must be proven to be effective! Studies of the implementation process are the result of an investigation into this failure of institutions to adopt a change in curriculum and failures of curricula to achieve the intended objectives. In a summary of research since 1971, Fullan and Pomfret (1977) listed fifteen studies on defining and measuring implementation. Also, an international seminar on curriculum implementation and its relationship to curricula development in science was held in Rehovot and Jerusalem, Israel in 1978 (Tamir, et al., 1979). These studies and others indicate a growing concern about the importance of understanding the implementation process of curriculum.

Implementation is defined as "the actual use of an innovation or what an innovation consists of in practice" (Fullan & Pomfret, 1977). An innovation can involve several dimensions of the educational system. Fullan and Pomfret (1977) outlined implementation as changes in: materials, organizational structure, role and behaviors, knowledge and understandings, and value internalization. Implementation of programs can occur through adoption or adaptation. Adoption refers to the decision to achieve a high fidelity of use, or close correspondence between the actual innovation processes in the classroom and those intended by the curriculum developers. Adaptation is a dynamic process of mutual change of the innovation and user to meet specific local needs (Tamir et al., 1979).

Fullan and Pomfret (1977) classified fifteen studies as either investigating implementation from the fidelity approach or the process approach. This classification system more clearly differentiates between the underlying philosophical basis of the various approaches to investigating implementation than do the categories of adoption and adaptation. Externally developed curriculum innovations enter a system through a process of adoption and are adapted to the local situation to varying degrees. Even curriculum changes that are part of an ongoing process of continued adaptation are often codified and formally adopted by the system. Studies using the fidelity approach investigate the degree to which the innovation in the local situation corresponds with the developers' intentions and assess the level of achievement of the predetermined objectives. The fidelity approach emphasizes evaluation of a well defined innovation. It is based upon the value of using large scale curriculum projects aimed at predefined outcomes to influence local education programs.

The process approach to investigating implementation emphasizes the dynamic nature of innovations. Changes in the curricula involve changes in the people, environment, materials, and other factors. Not all changes can be predicted or controlled. Not all outcomes can be predetermined. The process approach investigates the actual changes in the educational system that occur during the adaptation to an innovation. The emphasis is on description, analysis, and explanation of the dynamic interactions within the system. The focus is on explaining successful curriculum implementation efforts, which is based upon the value of capitalizing on unique situational characteristics. The process approach is consistent with recent trends in curriculum development indicating a shift toward greater involvement of local systems in planning and implementation (Tamir, et el., 1979).

The fidelity approach of implementation has been widely used in many large scale national curriculum projects. However, there have been many failures to find significant differences between traditionally defined experimental and control groups in evaluations of these projects. Three studies of the *Intermediate Science Curriculum Study (ISCS)*, using adoption as the criterion for use, found little difference among users and non-users of the program (Hefferman, 1973; Mann, 1972; Martinez-Perez, 1973). In a study by Hall and Loucks (1977) on Individually Guided Education (IGE) schools compared with non-IGE schools, they found no significant difference between groups on achievement. When the groups were assessed for degree of implementation of IGE, they found there was more within group variance of use of IGE processes than between groups. Hall and Loucks (1977) concluded that "if the variance associated with use/non-use goes unaccounted for in statistical analysis one result may be greatly increased error terms leading to inability to identify real differences." Perhaps the earlier, more traditional studies of implementation from fidelity approaches were too restrictive in their evaluation design and therefore failed to completely investigate the innovation in use.

A thorough evaluation of the implementation of an innovation by the fidelity approach would answer the following questions:

1) What is the correspondence between the dimensions of the innovation as defined by the program developers and the characteristics of the

curriculum as it is actually used?
2) What is the level of achievement of the predetermined objectives defined by the program developers for those users adopting the innovation?
3) Does maximizing the fidelity of the use of the innovations maximize the achievement of the predetermined outcome?

The first question is written to indicate the fidelity approach to measuring the degree of implementation of an innovation. The second question is included as a means to evaluate the success of the fidelity approach to implementation of achieving its goal of influencing the outcomes of educational systems. The third question is written to indicate the ability of the innovation to achieve the predetermined objectives if fully implemented as designed by the developers; and if increased fidelity of use will increase the level of achievement. A high positive correlation between fidelity of use and achievement of the desired outcomes lends support to the validity of the curriculum as a worthwhile instructional package.

Several studies have followed the process approach to implementation and have investigated the process of change within the users of the innovation. Hall, Wallage and Dossett (1973), found through studying the change process in implementation, that teachers undergo a series of developmental changes as they adjust to new roles and behaviors specified by an innovation. The *Levels of Use of the Innovation Interview/Rating System* (LoU) (Loucks, Newlove, & Hall, 1975) was developed in response to a need to find a method to adequately and efficiently measure the degree of implementation of an innovation. The LoU is a widely used instrument for meausring implementation. The LoU is used to measure the degree of change of the users involved in the process of adapting to the innovation.

The process approach to studying implementation investigates questions similar to those of the fidelity approach. The process approach investigates the following questions:

1) How have the users adapted to the innovation?
2) How are the users utilizing the innovation?
3) What are the outcomes relating to the present state of implementation of the innovation?

The first question describes the changes that have occurred within the user system in response to the innovation. These changes indicate the process approach identification of the degree of implementation. The second question is a description of the configuration of use of the innovation as it occurs in the actual classroom situations. The configuration is a documentation of the instructional process. The third question addresses the success of the innovation at achieving positive outcomes. The outcomes may not have been determined a priori and may even be unpredicted results from the innovation. Scriven's (1973) goal free evaluation procedures are consistent with the process approach philosophy of implementation.

PROBLEM

How valid are different approaches used to measure the degree of implementation of an innovative program?

PURPOSE

Recent evaluation studies have been performed which include measures of the degree of implementation of an innovation. Gene Hall et al. (1973) have devoted extensive efforts to the development and use of the *Concerns Based Adoption Model* of implementation. Very little attention has been focused on the validity of different approaches used to measure the degree of implementation. A study by Ellis (1979) investigated the relationship between the fidelity and process approaches of implementation. This study compared two different approaches to measurement, the LoU and the fidelity approach, respectively. The purpose of this study was to validate the *Degree of Implementation of an Innovation*. The innovation used in this study was the *Intermediate Science Curriculum Study* (ISCS) science program.

RESEARCH DESIGN

The construct validation procedure was used to examine the relationship among different methods of measuring the degree of implementation of an innovation. Three of the sets of variables in the study were used to investigate the construct validation. The level of use assigned by the LoU system, as a measure of the mutual adaptation of the user with the innovation, was one set. The fidelity approach measures were the second set and consisted of: (1) *Student Perception of the Use of ISCS* (SPU), a student questionnaire developed for this study to assess the degree of implementation of the program as defined by the developer, (2) *Teacher Interview of Use* (TIU), a teacher interview of use of the innovation developed for this study, and (3) *Document Analysis of Use* (DAU), an analysis of documents and equipment used in the program. The third set of variables consisted of: (1) the residualized achievement score (RESDACH), a measure of achievement controlling the score on the *Science Achievement Test* for the math subscore on the *Iowa Test of Basic Skills*, and (2) the *ISCS Attitude Questionnaire* (ATT), a measure of attitude toward science and problem solving.

Three of the hypotheses examined in the study were:

H_{o1}) No linear combination of variables exists of SPU, TIU, and DAU which results in a significant difference between groups of teachers categorized according to LoU;

H_{o2}) No relationship exists between linear composites formed from SPU, TIU, and DAU as one set of variables and RESDACH and ATT as the other set;

H_{o3}) No linear combination exists of the variables RESDACH and ATT which results in a significant difference between groups of teachers categorized according to LoU.

The rejection of the null hypotheses will lend support to the validation of the different approaches as measures of the proposed construct, the degree of implementation of an innovation. Null hypothesis one compares the fidelity approach measures with Levels of Use. Null hypotheses two and three examine the relationships of the two approaches to predicted student outcomes.

SAMPLE

The *Intermediate Science Curriculum Study* (ISCS) was the innovation used in this study. Fifteen ISCS teachers, representing all available ISCS teachers in a large midwest metropolitan city, were used in the study. An additional nine non-ISCS teachers volunteered for the study. Both groups were included to insure that there would be teachers at all degrees of implementation from non-use to extensive modification of use of the innovation.

Students enrolled in seventh grade science in ten different junior high schools were involved in the study. Seventh grade science is a required course; however, either ISCS or earth science met the requirement. The students involved in the study were drawn from different geographic locations and represented a cross section of students across the district.

Those classes involved in ISCS were utilizing Silver Burdett's revised version of ISCS' *The Natural World* (1975). The non-user category consisted of teachers involved in an earth science program for seventh grade students. The students and teachers participated on a voluntary basis to meet guidelines for research established by the district. There were 86 seventh grade classes and 1,730 students involved in the study.

INSTRUMENTATION

The LoU instrument was used as one approach to measuring the degree of implementation. The LoU instrument is an interview rating scale used to categorize subjects according to level of use of a pre-defined innovation. The data are gathered during a 30-minute focused interview. Assignment to categories is made upon analysis of the taped interview. Figure 1. displays the LoU category system. A subsample of one-third of the total taped interviews in the study was rated by a separate trained rater. Seventy-five percent agreement was reached for rating the tapes in the same category of use. One hundred percent agreement was reached between raters for rating the taped interviews as either the same category or adjacent category of use.

The three methods for measuring the fidelity approach were: (1) *Student Perception of Use of ISCS* (SPU), (2) *Teacher Interview of Use* (TIU), and (3) *Document Analysis of Use* (DAU). Items for all three methods focussed on the three dimensions of ISCS used in this study. These dimensions were: (1) Subject Matter/Materials, (2) Structure/Organization, and (3) Role/Behaviors. The teacher interview consisted of a fifteen-minute tape recorded interview session. The tape interview was rated on each of the three dimensions and assigned a value using the rating scale developed for this study.

Figure 1

LEVELS OF USE OF THE INNOVATION

LEVELS OF USE	DEFINITION OF USE
0 NON-USE	State in which the user has little or no knowledge of the innovation, no involvement with the innovation, and is doing nothing toward becoming involved.
I ORIENTATION	State in which the user has recently acquired or is acquiring information about the innovation and/or has recently explored or is exploring its value orientation and its demand upon user and user system.
II PREPARATION	State in which the user is preparing the first use of the innovation.
III MECHANICAL USE	State in which the user focuses most effort on the short-term, day-to-day use of the innovation with little time for reflection. Changes in use are made more to meet user needs than client needs. The user is primarily engaged in a stepwise attempt to master the tasks required to use the innovation, often resulting in disjointed and superficial use.
IVA ROUTINE	State in which use of the innovation is stabilized. Few if any changes are being made in ongoing use. Little preparation or thought is being given to improving innovation use or its consequences.
IVB REFINEMENT	State in which the user varies the use of the innovation to increase the impact on clients within immediate sphere of influence. Variations are based on knowledge of both short- and long-term consequences for clients.
V INTEGRATION	State in which the user is combining own efforts to use the innovation with related activities of colleagues to achieve a collective impact on clients within their common sphere of influence.
VI RENEWAL	State in which the user reevaluates the quality of use of the innovation, seeks major modifications of or alternatives to present innovation to achieve increased impact on clients, examines new developments in the field, and explores new goals for self and the system.

Excerpted from *The LoU Chart*, Procedures for Adopting Educational Innovations Project, Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, 1977.

The rating scale was developed by identifying the salient characteristics inherent to the ISCS program. Within each of the three dimensions four items were developed. The rater placed each class on a one to five continuum for each of the items in all dimensions. A rating of one on an item indicated that the class was entirely non-ISCS for that aspect of the dimension and a five indicated that the class was totally ISCS for that aspect.

The interrater reliability of the fidelity approach *ISCS Rating Scale* was established by correlating the ratings with an independent rater, for a subsample of one-third of the teachers. A 0.97 interrater reliability coefficient was established based upon 13 joint observations.

The student perceptions of the classroom educational process were gathered by the *Student Perceptions of Use of ISCS* questionnaire. Five students were selected at random from each class to respond to the questionnaire. There were ten items for each of the three dimensions, with responses to each item ranging from one to five on a Likert-like scale, corresponding to strongly agree to strongly disagree.

Actual in-class examination of materials and equipment in use was performed to gather data for the *Document Analysis of Use*. Textbooks in use, record books, equipment, supplementary materials, lesson plans, teacher grade book, and testing systems were examined to obtain descriptive data about the innovation in use. After complete examination of the materials, the class was rated using the rating scale developed for the three dimensions of ISCS.

An eighteen item *Science Achievement Test* selected from a composite of items developed and tested by the developers of ISCS was used in the study. The test measured the achievement of objectives for the first nine chapters of the text. A Kuder-Richardson — 20 internal consistency alpha value of 0.56 was derived for the student achievement test.

RESULTS

The central purpose of this investigation was to determine to what extent an innovative science program was implemented. Specifically the task was to perform a construct validation of the Degree of Implementation of an Innovation. This section will be organized under subheadings that correspond with the three null hypotheses used to test the construct validation.

Null hypothesis one (H_{01}) *Multiple Analysis of Variance (MANOVA)* was used to examine the relationship between LoU and the fidelity measures. It was predicted that different levels of use should have significantly different composite values for *Student Perception of Use of ISCS* (SPU), *Teacher Interview of Use* (TIU), and *Document Analysis of Use* (DAU). The LoU was collapsed into four categories to keep adequate cell sizes. The LoU categories were formed by combing LoU (O, I, and II) into category one, LoU (III) into category two, LoU (IVA) into category three, and LoU (IVB, V, and VI) into category four. These division points were chosen because category one represents all non-users, category two represents users adapting to the innovation, category three represents routine users, and category four represents users who are making improvement and changes in the innovation.

The users and non-users were pooled for the primary *MANOVA* analysis, with teachers used as the unit of analysis. The Wilk's multivariate test of significance was significant, *F* (3, 0, 9) = 53.88, $p<0.05$. One significant root was obtained, $p<0.05$, with an eigenvalue of 34.44 and canonical correlation of 0.99. Univariate *F* tests for *Student Perception of Use* (SPU), *Teacher Interview of Use* (TIU), and *Document Analysis of Use* (DAU) by LoU were all significant, $p<0.05$. Secondary analysis using the discriminant analysis procedure yielded one significant function as follows:

$$\text{LoU} = 0.005\ \text{SPU} - 0.70\ \text{TIU} - 0.40\ \text{DAU}.$$

Null hypothesis one (H_{01}) was therefore rejected, supporting concurrent validity of the Degree of Implementation of an Innovation construct.

An additional secondary *MANOVA* analysis, including only users in the calculation, found no significant difference between categories of higher levels of use for the composite fidelity measures. The primary relationship between LoU and the fidelity measures lies in their common ability to categorize teachers according to use and non-use of the innovation.

Null hypothesis two (H_{02}). The second procedure for the construct validation involved using canonical analysis to examine the relationship of the fidelity approach measures with predicted student outcomes. The two measures of student outcomes were the residual of the *Science Achievement Test* (RESDACH) and the *ISCS Attitude Questions* (ATT). The primary analysis included both users and non-users in the calculations. The classes were used as the unit of analysis for null hypothesis two (H_{02}).

Table 1 is presented to summarize the results of the canonical analysis for null hypothesis two (H_{02}). A significant relationship was found between the composite fidelity measures and the composite student outcome, $p<0.05$.

Table 1

SUMMARY OF CANONICAL ANALYSIS FOR NULL HYPOTHESIS TWO (H_{02})

n = 80

Number	Eigenvalue	Canonical Correlation	Wilk's LAMBDA	CHI Square	D.F	
1	.47	.68	.51	52.01	6	*
2	.05	.22	.95	3.71	2	n.s.

*p < .05

One significant function was derived as follows:

$$0.90\ \text{SPU} - 1.49\ \text{TIU and } 1.56\ \text{DAU} = 0.96\ \text{RESDACH} - 0.22\ \text{ATT}$$

Secondary canonical analysis for hypothesis two, using only users in the calculation, was significant, $p<0.05$. One significant function was derived with the same pattern of weights as in the overall analysis. Null hypothesis two (H_{02}) was therefore rejected and a significant relationship was found between the fidelity composite and predicted student outcomes.

Null hypothesis three (H_{03}). The third procedure for the construct validation involved using *MANOVA* to examine the predicted differences among categories of LoU for student outcomes. The primary analysis included both users and non-users in the calculations. The teachers were used as the unit of analysis. The eight categories of LoU were collapsed into the four categories used in the analysis of null hypothesis one.

There was a significant difference between categories of LoU for the composite student outcomes as indicated by Wilk's multivariate test of significance. $F(2, 0, 8) = 3.97$, $p<0.05$. One significant root was obtained, $p<0.05$, with an eigenvalue of 1.2 and a canonical correlation of .75. The univariate *F test results indicated that there was a significant difference between categories for residual student achievement but not for student attitudes. Secondary analysis using the discriminant analysis procedure yielded one significant function as follows:*

$$LoU = .98\ RESDACH - .19\ ATT$$

Null hypothesis three (H_{03}) *therefore was rejected lending concurrent validity to the construct.*

An additional secondary MANOVA analysis, including only users in the calculations, found no significant difference between LoU categories for higher levels of use for the composite student outcomes. The primary relationship between categories of LoU and predicted student outcomes lies in the differences between users and non-users of the innovation.

DISCUSSION

The results of this study lend support to the validity of the Degree of Implementation of an Innovation construct as a factor in the educational process. There is a strong relationship between the process approach and fidelity approach to defining the degree of implementation of an innovation. Both approaches to measuring the degree of implementation were significantly related to predicted outcomes of the program when including users and non-users of the innovation as the sample for the analysis.

The fidelity approach to measuirng the degree of implementation does differ from the process approach. When examining the relationship between the degree of implementation and predicted student outcomes, for only users of the innovation, the fidelity approach gives a strong relationship between higher fidelity of use and achievement. The process approaches does not exhibit a significant relationship between higher levels of use and achievement. This indicates that maximum attainment of predetermined objectives set by the curriculum developer is best achieved by following the fidelity approach to implementation.

The process approach to implementation emphasizes the mutual adaptation of the user and the innovation. With high levels of use, users modify the innovation to meet their own perceptions of the needs of their situation. This may lead to a strong individual program that maximally achieves many worthwhile outcomes. However, some of the outcomes may not have been predicted by the curriculum developers. Some of the predetermined objectives may not be emphasized by the user in the adaptation of the innovation. The result is that for users, as defined by the process approach, there is no significant relationship between student outcomes and higher levels of use.

When the purpose of the evaluation is to determine the degree to which a curriculum implementation has been successful at achieving predetermined objectives the fidelity approach is the valid and appropriate procedure to use. Curricular programs with predetermined objectives are the result of national, state, regional, and occasionally district efforts to provide instruction to meet the needs of the citizens served. These programs are adopted for use in individual classrooms within districts and often diverse variations of use can occur. When evaluating the overall effectiveness of the program at achieving the predetermined objectives, it is important to document all variations of use and additional outcomes; but, the central focus should be on the process outlined by the curriculum.

When the focus of the curriculum implementation is to facilitate the dynamic ongoing process of local curriculum adaptation and to emphasize the attainment of more global, open-ended goals, the process approach to implementation is appropriate. During the development and implementation phase of curriculum evaluation often teachers and districts go through a process of adapting to the innovation. Often program objectives are modified, or teaching processes changed in responses to informal and formal feedback from teachers and students. The success of the implementation phase would be enhanced by use of a quality formative evaluation design which would include appropriate information for making modification decisions. The LoU instrument provides a very effective model for explaining the teachers adaptation to the innovation. Including information on configuration of use in the LoU interview increases its utility as a measure of the actual teaching process. The LoU interview with a complete configuration of use is a good formative evaluation tool. Coupled with a goal free evaluation design, the process approach offers a method of evaluation of the "living" curriculum as it exists in schools in our society.

REFERENCES

Ellis, J. D. *The construct validation of the degree of implementation of an innovation.* Unpublished doctoral dissertation, The University of Kansas, 1979.

Fullan, M. & Pomfret, A. Research on curriculum and instruction implementation. *Review of Educational Research, 47* (2), 1977, 335-397.

Gross, N. & Herriott, R. *The dynamics of planned educational change*. Berkeley, California: McCutchan, 1979.

Hall, G. E., Wallace, Jr., R. C. & Dossett, W. A. *A developmental conceptualization of the adoption process within educational institutions. (Research on Concerns-Based Adoption Project, Working Paper 3006.) Unpublished manuscript. The Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, 1973. (ERIC ED 095 126.)*

Hall, G. E., & Loucks, S. A developmental model for determining when the treatment is actually implemented. American Education Research Journal, 1977, 39-49.

Hefferman, Daniel F. *A comparison of the effects of individualized science instruction with traditional science instruction in junior high school*. Unpublished doctoral dissertation, University of Rochester, Rochester, New York, 1973.

Loucks, S. F. *Levels of use of innovation: The conceptualization and measurement of a variable useful for assessing innovation implementation by individuals*. (Research on Concerns-Based Adoption Project, Working Paper 3003.) Unpublished manuscript, The Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, April 1977.

Loucks, S. F., Newlove, B. W. & Hall, G. E. *Measuring levels of use of the innovation: A manual for trainers, interviewers and raters.* (Research on Concerns/Based Adoption Project, Working Paper 3013.) Unpublished manuscript, The Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, 1975.

Mann, E. *A study of attitudes toward science of selected junior high school students after exposure to an individualized science program* (ISCS). Unpublished doctoral dissertation, Florida State University, 1972.

Martinez-Perez, L. A. *A study of self-concept, attitudes toward science, and achievement on a sample of seventh grade ISCS students versus seventh grade students in a non-individualized science class*. Unpublished doctoral dissertation, Florida State University, 1973.

Rowe, M. B. What research says. Elementary Science: A vanishing species. *Science and Children, 18* (1), 19-21.

Scriven, M. Goal free evaluation. In E. House (Ed.), *School Education*, Berkeley, California: McCutchan, 1973.

Tamir, P.; Blun, A.; Hofstain, A. & Sabar, N. *Curriculum implementation and its relationship to curriculum development in science*. Israel Science Teaching Center: Jerusalem, 1979.

Yager, Robert E. Crisis in science education. *Technical Report #21*. Science Education Center, University of Iowa, 1980.

Loucks, S. F. *An exploration of levels of use of an innovation and the relationship to student achievement*. (Research on Concerns-Based Adoption Project, Working Paper 3027.) Unpublished manuscript, The Research and Development Center for Teacher Education, The University of Texas at Austin, Austin, Texas, Spring 1976. (ERIC ED 123 805.)

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN STUDENT CHARACTERISTICS AND COGNITIVE GAINS OBTAINED BY UNDERGRADUATE STUDENTS IN AN AUDIO-VISUAL TUTORIAL SELF-PACED SCIENCE PROGRAM

ROBERT M. HOLLAND
MARY HARDIN-BAYLOR UNIVERSITY
BELTON, TEXAS

INTRODUCTION

One of the primary goals of education is to make the learning process more desirable and thus to affect learner attitudes toward the curricular format. Programmed instruction emerged as one alternative to traditional instruction during the 1950's with a variety of instructional techniques and teaching aids. Audio-visual tutorial instruction was one of the earliest techniques under the broad umbrella of "programmed instruction" and has been implemented and evaluated in a wide variety of educational settings. Science has seen its share of curricular materials used in this alternate format. Self-paced materials have provided students with the opportunity to study and learn at a rate congruent with their job and/or life style.

The modular format allows the curriculum developer the opportunity to use a wide variety of techniques and media and to optimize the instructional process for a wide range of student abilities and interests. Many educators believe that given sufficient time most students will reach an acceptable level of performance in a course, regardless of the format. Even so, the curricular format could be very important in terms of attitudes toward the subject or toward the mode of instruction.

PROBLEM

What effect does an audio-visual tutorial self-paced program have on students' cognitive gains and attitude toward the program?

PURPOSE

The purpose of this investigation was to evaluate both cognitive knowledge gained and attitude change as a result of exposure to four senior level biophysical analysis audio-visual self-paced modules. In addition, attention was focused on identifying undergraduate student characteristics which correlated positively with gain scores and attitude changes. The four modules along with appropriate activities and pretests-posttests were developed and then utilized with the class. Topical areas represented in the four modules were: 1) Computer Programming (Fortran IV), 2) Elementary Electronics, 3) General Spectroscopy, and 4) UV-Visible Spectroscopy.

RESEARCH QUESTIONS

The following research questions were investigated:

1. What are the cognitive gain scores for each student taking the audio-visual tutorial self-paced modules?
2. Is there a positive correlation between cognitive gain scores in each of the four modules and student characteristics?
3. Is there a positive correlation between attitude changes and cognitive gains for each of the four modules used in this evaluative study?
4. Did prior experience with a non-traditional format of instruction correlate with changes in attitude toward the audio-visual tutorial self-paced mode of instruction used for presentation of these modules?
5. Are certain characteristics positively correlated with a change in attitude toward an audio-visual tutorial self-paced mode of instruction?

Student characteristics used for assessment in questions two and five are: 1) Grade Point Average (GPA) scores, 2) Scholastic Aptitude Test (SAT) scores, 3) prior experience with self-paced non-traditional type courses, and 4) total semester hours in computer science, mathematics, biology, chemistry, and physics.

RESEARCH DESIGN

The investigation basically involved two major tasks: 1) development of all questionnaires, modular materials and evaluative instruments and, 2) the actual testing of the students. The basic design of the study was a one group pretest-posttest arrangement.

SAMPLE

The subjects of this study were fairly homogenous in terms of academic and professional goals. Most were junior or senior in classification, majoring in the life-sciences, and planning careers in medicine, dentistry, and health science related vocations. There were 28 students who signed up for and participated in this study within the framework of Zoology 356, Biophysical Analysis.

PROCEDURE

The development of the instructional materials and the preliminary evaluation took place during the fall semester at a large south central Texas university. In the interest of clarity the overall procedure is best illustrated by the following stages: 1) development of a questionnaire and attitude survey, 2) development of performance objectives for each module, 3) development of modular material (textual, visual, activities, pre- and posttests), 4) production and programming of the module, 5) trial with students, and 6) evaluation as detailed in the study.

INSTRUMENTATION

Basically four instruments were developed for utilization with this investigation. They are as follows:

1) general questionnaire (given at the beginning of the course).
2) attitude survey, both pretest and posttest (given before Module One and at the conclusion of Module Four).
3) module pretest (preceding each of the four modules).
4) module posttest (given after each module is concluded).

The general questionnaire allowed research questions two and five dealing with student characteristics to be assessed in some detail. Questions such as subject matter, social, academic, non-traditional instructional and study habit backgrounds were called for to see if any correlation existed either in a positive or negative sense as they related to attitude or academic gains where the modules were concerned. The attitude survey was used as both a pretest and posttest to assess an overall change in attitude and to in turn see if this change correlated positively or negatively with cognitive gains registered on the modules. The non-traditional mode of instruction as it relates to the attitude changes was also examined.

The attitude survey used a Likert-scale format with statements concerning audio-visual self-paced instruction ranging from 1 (strongly agree) to 5 (strongly disagree). The pretest-posttest questions were developed from specific objectives developed for each module. The questions were subjected to both validity and reliability tests and questions were in turn placed in two parallel test forms. The general questionnaire, in addition to the specific information about student background, provided information which could be used in future studies.

HYPOTHESES

The first question of this investigation is stated in a linear fashion with the remaining four stated in the null hypothesis (H_o) format. They are as follows:

1. What is the cognitive gain score for each student after completion of the respective modules?
2. No relationship exists between cognitive gain scores for each of the four modules and the student characteristics.
3. No relationship exists between a change in attitude toward the audio-visual self-paced mode of instruction and cognitive gain scores.
4. No relationship exists between a change in attitude toward the audio-visual self-paced mode of instruction and prior experience with a non-traditional format of instruction.
5. No relationship exists between a change in attitude toward an audio-visual tutorial self-paced mode of instruction and student characteristics.

The first question calls for gain scores for each student or each module. These results are used in other questions where cognitive gain scores are required. The five hypotheses were tested and accepted or rejected at the 0.05 confidence level.

DATA ANALYSIS

In order to test hypothesis one, student gain score means on all four modules were subjected to a t-test (a modification of Fisher's t) for significant gains. SPSS (Statistical Programs for the Social Sciences) subprogram, NONPAR CORR., which measures the strength of relationship between two interval-level variables was used to analyze the data for possible correlations. This nonparametric statistical technique was used since the level of measurement indicated by the data was either nominal or ordinal. The statistical program has sufficient power to be useful with all of these levels of measurement for the remaining four hypotheses.

RESULTS

The first research hypothesis is stated in a question form since it deals only with the cognitive gain scores for each student on each module (see Table 1).

Table 1

STUDENT POSTTEST MEAN GAIN SCORES FOR THE FOUR MODULES

MODULES:	(1) Computer Program	(2) Electronics	(3) General Spectrocopy	(4) UV-VIS
	3.3	1.8	1.2	2.3

When comparing pretest and posttest mean gain scores of the students on the four modules none of the t-tests approached significance at the 0.03 level. Thus there was no significant difference between the pretest and posttest mean gain scores.

The second research hypothesis stated there was no relationship between mean gain scores for each of the modules and student characteristics. The characteristics used were GPA score, SAT score, prior experience with self-paced non-traditional type courses and prior experience with computer science, and total semester hours of mathematics and science courses. Following are the results of the analysis:

1. There were no significant correlations between mean gain scores on the modules and GPA.
2. There were no significant correlations between mean gain scores and SAT scores.
3. Mean gain scores were compared with seven types of self-paced non-traditional type courses.

Significant correlations between mean gain scores on the Computer Programming Module and CBE mean gain scores on the UV-VIS Module and PSI were found.

Table 2

CORRELATION BETWEEN COGNITIVE GAIN SCORES AND PRIOR EXPERIENCE WITH SELF-PACED NON-TRADITIONAL TYPE COURSES

MODULE		PSI**	CBE	CAI	AV	MOD	SELF	NON
Computer Programming	Correlation	−.0145	.3721	−.0092	−.0697	.0513	−.1270	−.0484
	Significance Level	.471	* .026	.482	.363	.398	.260	.404
Electronics	Correlation	−.1661	−.2425	−.1646	.0173	−.2936	−.1489	−.0963
	Significance Level	.200	.107	.202	.466	.065	.225	.314
General Spectroscopy	Correlation	−.0216	−.0891	−.2692	.2509	.0382	.1441	.1081
	Significance Level	.457	.326	.084	.099	.424	.233	.293
UV-VIS	Correlation	−.3990	−.0962	−.0827	−.0436	−.0834	.2177	.2418
	Significance Level	* .018	.314	.338	.413	.337	.133	.108

*p 0.05

**PSI — Personalized System of Instruction
CBE — Computer Based Education
CAI — Computer Assisted
AV — Audio-Visual Format
MOD — Module or Minicourse Format
SELF — Self-Paced Course Format
NON — Other Non-Traditional Format

Table 3

CORRELATION BETWEEN COGNITIVE GAIN SCORES AND TOTAL SEMESTER HOURS IN COMPUTER SCIENCE, MATHEMATICS, BIOLOGY, CHEMISTRY, AND PHYSICS

MODULE		COMPUTER SCIENCE	MATHEMATICS	CHEMISTRY	PHYSICS	BIOLOGY
Computer Programming	Correlation	−.3661	.1950	.1234	−.0122	.1895
	Significance Level	* .028	.161	.266	.476	.168
Electronics	Correlation	.1037	−.2476	.0587	.2697	−.0885
	Significance Level	.300	.102	.384	.083	.328
General Spectroscopy	Correlation	.0613	−.0451	−.0572	.1104	.0003
	Significance Level	.379	.410	.387	.288	.500
UV-VIS	Correlation	.2992	−.1098	.2146	.0816	.3494
	Significance Level	.061	.290	.137	.340	* .035

No other significant correlations between mean gain scores on the modules and other self-paced courses (see Table 2).

There were two significant correlations between mean gain scores on two of the modules and science semester hours. Overall there were not enough significant correlations and the second research was not rejected (see Table 3).

The third research hypothesis stated that no relationship exists between a change in attitude toward audio-visual self-paced instruction and cognitive mean gain scores. Results of analysis revealed no significant correlation between mean gain scores and the AV self-paced instructional mode. The null hypothesis was not rejected.

The fourth research hypothesis stated that no relationship exists between change in attitude toward audio-visual self-paced instruction and prior experience with a non-traditional format of instruction. No significant relationships were found except between attitudes and the Personalized Self-Instructional method and A-V experiences. Therefore, the null hypothesis was not rejected.

Question five stated that no relationship exists between change in attitude toward audio-visual tutorial self-paced instruction and student characteristics. There were no significant relationships between attitudes and student characteristics. Therefore the null hypothesis was not rejected.

DISCUSSION

This investigation centered around the preliminary evaluation of four audio-visual tutorial (A-V-T) self-paced instructional modules. While there were no significant results or correlations found, no negative results were identified. Thus it can be concluded that an A-V-T self-paced mode can be a viable alternative to traditional instruction methods. It does not hinder the learning of abstract ideas presented in the modules studied in this investigation.

The limited number of students, lack of a control group for comparison purposes, and perhaps the specific nature of the science materials used may have mitigated against clear results supporting the A-V-T format. A review of the research literature where limitations such as those previously identified did not reveal conclusive evidence supporting the alternate instructional format either (see Holland, 1977). Comparative studies seldom prove one pedogogical technique superior to another but rather allow the researcher an option to demonstrate the effectiveness of an alternate method.

Student reaction, feedback and gain scores overall showed that the A-V-T learning mode is a successful instructional tool for presenting science content. The reactions were positive in terms of visual appeal and the senses used to interact with this learning mode. The coupling of the audio with distinct visuals was received favorably by the students in the study. The activities were judged to provide a very good means of self-assessment for the students in the course. The instructional quality could be further enhanced with activities and other instructional materials.

In addition to the research hypotheses tested and the results recorded above, a questionnaire was distributed at the end of each module which asked for student reaction to the module under study. A summary of these reactions for reader perusal follows:

- The modules were considered to be lengthy by some
- The modules took an excessive amount of outside time when computer activities were required
- Progress through the materials was uneven
- The activities received a favorable and positive rating
- Students would have preferred more textual material and handouts to supplement the A-V presentation
- The lesson caused the student to be frustrated at times due to complexity of the material or activity
- Written explanations for the previous statement indicated that frustration was due to complex computer activities and inability to stop the A-V program and ask questions on the spot
- Using a difficulty rating scale from zero to ten, with ten being the most difficult, the rating scale average was around seven for the four modules
- Students preferred blue or purple slide backgrounds rather than red
- Placing the instructional components, in terms of helpfulness on a one-through-five scale, students rated discussion sessions as the least helpful on all four modules.
- This is perhaps due to the fact that such sessions were held only once a week. The most helpful component was the audio tape for the electronics and general spectroscopy modules, while textual material for the UV-VIS module and activities for the computer programming module were rated as most helpful.

IMPLICATIONS

Several of the more prominent contributions of this investigations are:

- Students can use an A-V-T self-paced instructional format, for highly technical subject matter, learn and apply this knowledge to a wide range of activities
- The development of these materials have been shown to be workable and serve as models for future development
- Students realized the specific contributions of the computer as it relates to these instrumental techniques
- The investigation results add to existing knowledge which shows the feasibility of presenting materials by an A-V-T self-paced instructional format.

The contributions listed above add to and emphasize ideas already put forth in previous studies (see Holland, 1977). These are valid contributions since educational research points out that curricular innovations and acceptance of non-traditional curricular formats require input from a wide range of disciplines and learning environments in order to gain wide acceptance. Many revisions and diverse student populations will, in time, result in clearer trends.

The advantages of being able to revise and improve the A-V-T modules by obtaining student feedback and reactions to the instructional mode and providing for individual differences in terms of content, rate of learning, and variety of sensory input are strong reasons for using this pedagogical approach. James D. Russell stated it quite succinctly when he wrote, "the basic question is, *Who* learns *What* under *Which Conditions* and in *How Much Time?* (p. 86)." Systematic programs of instructions of which A-V-T self-paced instruction is one, help to answer the above question and add to the repertoire of learning approaches and techniques for science instruction.

REFERENCES

Corey, S. M. The nature of instruction. Chapter 1, *NSSE Yearbook*, University of Chicago Press, 1967.

Dale, Edgar. Historical setting of programmed instruction. Chapter 2, *NSSE Yearbook*, University of Chicago Press, 1967.

Dwyer, Francis M. *A guide for improving visualized instruction.* Learning Services, State College, Pennsylvania, 1972.

Holland, R. M. *Development and preliminary evaluation of audio-visual tutorial modules in biophysical analysis.* Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1977.

McDuffie, Jr., Thomas E. Relationships between selected personal characteristics and achievement, attitude and success on an audio-tutorial biology program. Doctoral dissertation, Temple University, *Dissertation Abstracts*, 1973, *XXXIV* (4), 1729a-30a.

Mintzes, Joel J. A comparative study of student characteristics as correlates of performance in a college biology program employing audio-tutorial teaching modules. Doctoral dissertation, Northwestern University *Dissertation Abstracts*, 1974, *XXXV* (6), 3543a-4a.

Mintzes, Joel J. The A-T approach 14 years later — A review of recent research. *Journal of College Science Teaching*, IV (4), March 1975, 247-52.

Postlethwait, S. N.; Novak, J. & Murray, H. *An integrated experience approach to learning.* 2nd Edition, Minneapolis, Minnesota: Burgess Publishing Company, 1969.

Russell, James D. *Modular instruction: A guide to the design, selection, utilization and evaluation of modular materials.* Minneapolis, Minnesota: Burgess Publishing Company, 1974.

DETECTING AND CLASSIFYING THE ORIGINS OF SCIENCE MISCONCEPTIONS

RICHARD JAMES TREMBATH
CHISHOLM INSTITUTE OF TECHNOLOGY
FRANKSTON, VICTORIA
AUSTRALIA

INTRODUCTION

The existence of people who believe particular science ideas which are really misconceptions has been documented for a long time. Possession of scientific misconceptions may be and can be a handicap to citizenry required to make judgements about critical science issues. For instance, should a new nuclear power plant be built, or should public lands be leased for mineral and energy exploration, or should a certain part of the state be used for a nuclear waste dump site, or should there be increased fundings for medical or technological research efforts? Certainly misconceptions related to any or all of these important questions might result in decisions that could be detrimental to all concerned.

PROBLEM

It can be said that we currently live in a society that is kept informed through such mass media as television, magazines, newspapers and radio. We are constantly being bombarded with information whether we desire it or not. We receive the latest information including science discoveries before the ink is dry on the written page. So do high school and university students in the midst of the information explosion possess scientific misconceptions? If they do what is the nature and origin of these misconceptions?

PURPOSE

There appears to be some evidence that high school and university students believe that certain science misconceptions (e.g., sound will travel through a vacuum) are correct (cf. Bailey, 1963; Boyd, 1966; Hancock, 1940; Rogers, 1961; and Trembath, 1980). Since 1887 educational researchers have compiled lists of scientific misconceptions and have tested groups of students to find what percentage of them possess particular misconceptions. Little of this research, however, has gone beyond the stage of surveying the frequency of misconceptions. But no systematic determination of the origins of misconceptions has been reported in the literature and apparently no attempt has been made since Bacon in 1620 to classify misconceptions on the basis of their origins.

The main purpose of this study was to:

(a) construct an instrument to identify scientific misconceptions based on current scientific thought, possessed by high school and university students;

(b) identify the origins of scientific misconceptions;

(c) classify the origins of scientific misconceptions.

RESEARCH QUESTIONS

1. Can a reliable and valid test of scientific misconceptions be constructed which (a) samples questions from a wide range of sciences, (b) requires responses at and above the cognitive level of recall of knowledge, and (c) identifies sufficient misconceptions in high school and university students to allow an investigation of the origins of scientific misconceptions?

2. Does the time-consuming clinical method of investigation yield more information about the origins of misconceptions than a written method of investigation?

3. Does the use of the clinical method of investigation elicit enough information from those who possess misconceptions to reveal the manner in which they originate?

4. Are there any correlations between the number of misconceptions possessed by students and factors such as the number of science courses taken?

5. Are there different types of scientific misconceptions which can be classified on the manner in which they originate?

6. Because the investigator teachers in an Australian university, interest was focused on the similarities and differences of scientific misconceptions held by students in the two countries. Do American students from a central Texas city and students from Melbourne, Victoria, Australia exhibit similar misconceptions?

RESEARCH DESIGN

SAMPLE

A pilot test of science misconceptions was developed and administered to 64 (grade 9) randomly selected high school students in a randomly selected high school. All testing was conducted in a medium sized school system (57,000 students) in a central Texas city. From this a main test was constructed. This second test was administered to 47 randomly selected students at a major Texas university and 66 students randomly selected from an Australian university.

PROCEDURE

A selection of scientific misconceptions was made from those cited in the research literature, from newspapers, magazines, radio and television, and from personal experience as a science teacher. These misconceptions were formed into a pilot test. Two forms of the test were constructed. Both were identical except that in form A the questions were true-false-don't know (13) and multiple choice (8). In form B the students were asked to write the reasons why they chose the answers they did. After form A was administered the students were interviewed individually in order to obtain their reasons for choosing each answer.

Based on frequency rates in which only questions that received high frequency counts were included on the main form identified as the Trembath Test of Scientific Misconceptions (TTSM).

Next the TTSM was administered to 47 university students at a large major

Texas university. The TTSM contained 16 true-false-don't know items, 15 multiple choice items, and six short-answer items. This required 15 minutes. Then each student was interviewed individually for 15-30 minutes to find out why they chose each answer. All responses were recorded for later evaluation and classification.

The same procedure was carried out in an Australian university with a total of 67 students. No steps were omitted and all responses were recorded for later evaluation and classification.

Five science educators were asked to rate the TTSM on a five point Likert scale. The scale was to measure the importance of the misconceptions to the needs of a knowledgeable and literate citizenry. Below is a summary of the

Table 1

PRINCIPAL MISCONCEPTIONS TESTED IN EACH OF THE ITEMS IN SECTION B OF THE TTSM

Item	Principal misconception being tested
1	Animals breathe out pure carbon dioxide
2	The temperature of a liquid can be raised above its boiling point by increasing the heat applied to the liquid
3	Weight is reduced when burning takes place
4	Fish chemically decompose water to obtain oxygen
5	Moving to daylight-saving time changes the length of daylight
6	Freezing kills bacteria
7	The force of gravity is very much less on the top of a high mountain
8	In a collision, a force pushes passengers in a car forward (i.e., the reverse of Newton's First Law of Motion)
9	A heavy object accelerates at a greater rate than a lighter object
10	The hotter weather in summer is caused by the Earth moving closer to the Sun
11	Revolution of the Earth around the Sun is the cause of sunrise and sunset
12	Tides on Earth are caused by the Moon only
13	Most of the substance from which a tree is made comes from dissolved matter in the soil
14	The phases of the Moon are caused by the shadow of the Earth on the moon
15	If an object floats in water it must be made of a substance less dense than water

misconceptions used in the TTSM (Table 1 and 2). Face validity was in the range of 1.6-2.4 (important) for 34 of the 37 items. This suggests that the TTSM is a valid measure of importance.

Table 2

THE PRINCIPAL MISCONCEPTIONS TESTED BY ITEMS IN SECTION C OF THE TTSM

Item	Principal misconception being tested
1	A jet aircraft is driven forward by gases from the rear of the engine pushing on the air
2	Food remains cold in an icebox because cold flows from the ice
3	The inheritance of a long neck in giraffes is due to the inheritance of acquired characteristics
4	The two most common gases in the air are hydrogen and oxygen
5	Photosynthesis manufactures only one substance (generally thought to be sugar, oxygen, or hydrogen)
6	An insect is a bug (or some other inadequate description)

In order to ascertain the validity of the instrument used, content validity was deemed appropriate (Ary, Jacobs, & Razavich, 1972). It was further decided that scientific misconceptions be sampled from biology, physics, chemistry, health science, astronomy, and meteorology for adequate sampling validity. And finally, questions were designed to require two different intellectual abilities to answer them. Type A questions were those which required simple recall abilities and type B were those which required higher intellectual skills to answer them (e.g., comprehension, analysis, synthesis). The latter type questions were introduced in this study in an attempt to find some of the ways in which students acquire misconceptions which are not due simply to recall. This method breaks new ground in introducing those questions which require higher level mental activity for their solution (i.e., these questions required logical thought).

Using the Kuder-Richardson 20 formula reliability for the TTSM was found to be 0.68. While some may view this as somewhat low Nunnally (1967) believes that "in the early stage of development research . . . one saves time and energy by working with instruments that have only modest reliability, for which purpose reliabilities of 0.60 and 0.50 will suffice (p. 226)." By this standard, the reliability of the TTSM is satisfactory.

RESULTS

A 37-item test of science misconceptions was devised which sampled from six science areas and required responses above that of recall of knowledge for 19 of the items. The average number of correct answers scored by individuals was 13.8 (S.D. = 5.6) with a range of 2-24.

In order to determine if a valid test of scientific misconceptions could be constructed, 47 students took the 37-item TTSM. The average number of scientific misconceptions was 16.1 (S.D. = 3.1) with a range of 9-24. This misconception rate was high enough to provide sufficient material for clinical interviews.

The clinical method of investigation enabled the origins of 99% of the scientific misconceptions to be determined whereas the written method allowed the origins of only 41% of misconceptions to be determined. The difference between these results is statistically significant at the 0.0001 level of confidence.

Forty-seven students took the TTSM and 36 (77%) were interviewed individually. The interviews yielded 482 protocols. Except for 11 items of this total, interviews revealed the origin of each scientific misconception. Thus it was possible to detect the origin of the scientific misconceptions in 97.7% of the cases identified.

It was found that the amount of science taken at grade 10 and the number of scientific misconceptions believed is related and significant at the 0.05 level. However there appears to be no correlations between the amount of science taken in grades 11, 12, college level and number of scientific misconceptions.

The purpose for devising a classification system for the origins of scientific misconceptions was to demonstrate that there are many ways in which they arise and are acquired. The classification system which was devised was based on a careful analysis of the misconceptions recorded. These origins of the

Table 3

FREQUENCIES OF TYPES OF ORIGINS OF MISCONCEPTIONS

Type of origin of misconception	Frequency of type
Quasi misconceptions	
1. Guessed answer	15
2. Misunderstood question while taking test	38
3. Understood question during test, realized answer during interview	0
True misconceptions	
4. Incorrect knowledge of fact being tested	259
5. Incorrect knowledge of subordinate of concept being tested	18
6. Made incorrect logical step	141

misconceptions were divided into six main types and 27 sub-types. A summary of the types and the frequency of responses is shown in Table 3.

American university students' performance on the TTSM was compared with Australian university students' performance to see if there were any differences which might be due to science courses. Results are summarized in Table 4.

Table 4

MISCONCEPTIONS EXHIBITED ON THE TTSM BY EACH STUDENT IN THE AMERICAN AND AUSTRALIAN SAMPLES

	American (n = 47)	Australian (n = 66)
Mean	16.9*	15.12*
Standard Deviation	3.13	4.06

*t = 1.36
not significant

DISCUSSION

A principal aim of this investigation was to identify the ways in which students acquire scientific misconceptions. To gain this information, test instruments had to be designed to detect these misconceptions because tests used in previous scientific misconception studies had many faults. For an extended discussion see Trembath (1980).

The test instruments cited in some earlier studies of misconceptions contained a large proportion of items which were superstitions or folklore. Such items are of limited interest to the present study as they are presumably acquired by contact with other people. The manner of acquisition was described by Maller and Lundeen (1933). The present study was mainly concerned with those misconceptions of general scientific principles and facts which might be taught in schools and where, it would be hoped, incorrect information was not deliberately taught. These scientific misconceptions often reveal irrational or illogical thinking. For the present study, a test was designed which contained misconceptions of this type.

The Pilot Tests used in this study were designed to reveal some of the misconceptions which students possessed. After students had taken the test, an attempt was made to find the manner in which the misconceptions were acquired. Two methods for acquiring this information were used — a personal interview between the research and each student (Pilot Test, Form A) and a written response on the test paper (Pilot Test, Form B).

In personal interviews, the interviewer used the clinical method to try to establish the manner of acquisition of each misconception which was revealed to the Pilot Test, Form A. The interviews were tape recorded and the interviewer kept a written record of the interview on a recording sheet designed for the purpose.

The purpose of the two forms of the test was to find out whether a difference existed in the amount of information collected by the personal interview form of the test or the written-response form. This was a particularly important investigation because the personal interviews demanded a great deal of the test. Furthermore, the personal interview program was more disruptive of the school programs than the written-response program. Unless it could be shown that the interview program gave much more information on the origin of misconceptions there would be little point in pursuing this interview program.

The clinical interview conducted as part of the Form A of the Pilot Test revealed the origins of 99% of the misconceptions displayed but the written method used in Form B revealed the origins of only 41% of the science misconceptions.

These results show quite convincingly that the time-consuming clinical interviews generated far more information about the origins of misconceptions than the faster-to-administer items requesting a written response after each item.

Consequently, the Main Test (TTSM) was designed on the assumption that a clinical interview would be associated with it. The test items were selected from areas of physics, chemistry, health, astronomy and meteorology and required answers at the cognitive levels of recall, comprehension and application.

Five judges rated each item on a five-point Likert scale. The rating of around "important" for 34 of the 37 items suggests that the TTSM is a valid measure of the importance of the need for a scientifically literate society.

In a study like the present one, it is tempting to seek to link the number of misconceptions a student possesses with all sort of variables. The present study sought to find correlations between the number of misconceptions believed by a student and such factors as the number of science courses taken. It is easy to understand that there may be a link between number of semesters of school and college science taken by a student.

The significant positive correlation between the number of semesters of science taken at Grade 10 level and the number of misconceptions believed is surprising — the more science taken, the greater the number of misconceptions. A possible explanation of this correlation could be that material is taught at Grade 10 which many students find confusing. This is mere speculation and a further study is needed to, firstly, check that this correlation exists and, secondly, to search for any casual relationships if the correlation is substantiated.

The purpose of devising a classification of the origins of misconceptions was to show that there are many ways in which misconceptions are acquired. If this study is to affect teaching practice, it is essential that teachers realize that misconceptions can arise in many ways. This classification reveals those ways and points the way to improvement of teaching procedures.

The classification system devised for this study was built on the classifications devised by Bacon (1620/1944); Conklin (1919), Maller & Landeen (1933); Ausubel (1963, 1969); and Kuethe (1963).

All of these earlier classifications are lacking in sufficient detail for teachers and researchers to be able to identify and make a useful classification of the errors that students are making. The present classification shows the errors that students were making when they acquired these misconceptions.

One of the values of the new classification system is that when teachers are dealing with topics in which misconceptions commonly arise, they know in advance, not only that these misconceptions will arise, but the likely manner in which they will arise. Such knowledge could make a significant improvement to the teaching in these traditionally difficult areas.

As was noted previously a major purpose of this study was to devise a classification system for the origins of misconceptions which could be used reliably by trained judges. If trained judges were unable to agree on the classification of the origins of individual misconceptions revealed in interview protocols, then the classification is useless.

After the researcher had devised the system of classification for the origins, three science educators were trained in the use of the classification system and then asked to classify the origins of 10 misconceptions chosen at random using the Trembath Classification of the Origins of Scientific Misconceptions (TCOSM). Of the 10 randomly chosen misconceptions, the origins of 9 were independently given an identical classification and the tenth was independently given a slightly different classification by each judge. Agreement about the tenth origin was reached when the protocol was discussed. Thus, by use of the TCOSM, different judges can classify the same set of misconceptions and obtain classifications which do differ significantly from each other. The classification system is a valid one for classifying scientific misconceptions. For an expanded discussion of the system see Trembath (1980).

The method of classifying the origins was thus reliable when used by a trained judge. Without this reliability, much of the power of this investigation would be lost.

Now that the classification system seems to be a reliable and useful instrument for researchers and teachers to use, there should be a strategy devised to ensure that misconceptions about scientific topics are identified and that this knowledge is utilized to improve standards of education.

The study was extended to show that the misconceptions which were discerned in students in a central Texas city were not a chance phenomenon but could be replicated in another part of the world in a different (although not radically different) English-speaking culture. The point being made here is that similar misconceptions occur in a culture where different text books are used, where teachers are educated and administered differently and where school systems differ significantly.

The performance of the two groups on the TTSM was not significantly different, as shown by the t-test for a difference between two independent means. That two groups of students in different cultures did exhibit virtually the same numbers of misconceptions on a test suggests that the acquisition of these misconceptions was not due to local factors (e.g., school curricula, textbooks used in schools, education of teachers) or cultural factors unique to each society (e.g., levels of education attained by citizens; books read, and television watched, by citizens), but seems to be due to more general factors.

The manner in which these misconceptions were acquired reveals a suggestion for their probable universality — the scientific misconceptions were often acquired by students making unjustified assumptions and by students making errors in logic. It is easy to believe that these errors could be widespread among humankind.

The similarity of performance of the American and Australian groups on the TTSM suggests only that the possession of misconceptions may be a widespread phenomenon. In a previous study, Za 'Rour (1975) found misconceptions amongst students in Israel, but he made no direct comparison with students in any other country.

IMPLICATIONS

This study opens up large new areas of investigation which have scarcely been touched by earlier research. These areas have immediate practical application to any teaching/learning situation. Listed below are some of the research investigations which flow from the present study:

a) A classification should be made of classifications in all subject areas;
b) Does the possession of a misconception hinder subsequent sequential learning?
c) Do teachers correctly perceive the scientific misconceptions possessed by their students?;
d) Can teaching techniques be modified to reduce the frequency with which scientific misconceptions arise in students?;
e) Is the possession of scientific misconceptions a geographically widespread phenomenon?

REFERENCES

Ary, D., Jacobs, L. C., & Razavich, A. *Introduction to research in education.* New York: Holt, Rinehart & Winston, 1972.

Ausebel, D. P. Cognitive structure and the facilitation of meaningful verbal learning. *Journal of Teacher Education*, 1963, *24*, 217-30.

Ausubel, D. P. & Robinson, F. G. *School learning.* New York: Holt, Rinehart & Winston, 1969.

Bacon, F. *Novum organum.* New York: Willey Book Co., 1944.

Bailey, D. E. *An analysis of science misconceptions held by secondary pupils in North Carolina.* Unpublished doctoral dissertation, University of North Carolina, 1963. Dissertation Abstracts International, 1963. University Microfilms 63 3542.

Boyd, C. A. A study of unfounded beliefs. *Science Education*, 1966, *50*, 396-8.

Conant, J. B. *Science and common sense.* New Haven: Yale University Press, 1962.

Conklin, E. S. Superstitious belief and practice among college students. *American Journal of Psychology*, 1919, *30*, 83-102.

Hancock, C. H. An evaluation of certain popular science misconceptions. *Science Education*, 1940, *24*, 208-13.

Kuethe, J. L. Science concepts: A study of "sophisticated errors." *Science Education*, 1963, *47*, 361-4.

Maller, J. B. & Lundeen, G. E. Sources of superstitious beliefs. *Journal of Educational Research*, 1933, *26*, 321-43.

Milkent, M. M. Personal communication, Nov. 1978.

Rogers, W. P. What do children believe? *Science Education*, 1961, *45*, 131-7.

Trembath, R. J. *The frequencies and origins of scientific misconceptions*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, 1980.

Za 'Rour, G. I. Scientific misconceptions among certain groups of students in Lebanon. *Journal of Research in Science Teaching*, 1975, *12*, 385-91.

7

THE EFFECTS OF PARTICIPATION IN SELECTED ACTIVITIES FROM THE BIOLOGICAL SCIENCE CURRICULUM STUDY-ELEMENTARY SCHOOL SCIENCES PROGRAM ON THE ENHANCEMENT OF LISTENING SKILLS

JENNIFER R. W. SWIFT
AUSTIN INDEPENDENT SCHOOL DISTRICT
AUSTIN, TEXAS

INTRODUCTION

Educators, concerned about the lack of science teaching in the primary grades, have addressed themselves to the question: Why should young children experience science? Somewhat surprisingly, this has led to a focusing of research on the effects of participation in inquiry oriented science programs on basic language skills (reading readiness and oral communication skills) that underlie general academic success (Ayers and Mason, 1966; Ritz, 1969; Neuman, 1974; Huff and Languis, 1973; Bethel, 1974; Rowe, 1968, 1975). Similarly, this study was designed to investigate the effects of a science program at the elementary level, the *Biological Sciences Curriculum Study-Elementary School Sciences Program (BSCS-ESSP)*, on the enhancement of listening or receptive language skills.

The orientation of many elementary school programs developed in the 1960's is that young children should experience science in a way analogous to that of scientists. Allowing children to experience science directly has led to the delineation of so-called science process skills that underlie the investigative process used by scientists. These skills include observing, identifying, naming, classifying, predicting, inferring and communicating (Butts, 1973). Such skills bear a close resemblance to those identified by reading and language arts investigators (Guszak, 1970; Duker, 1966; Phillips, Dunham, Brubaker, & Butt, 1970; Stauffer, 1970) as components in producing and decoding oral and written communication. An avenue of research developed which focused on the enhancement of reading readiness and language arts skills of those children participating in programs that are activity-centered and inquiry-oriented.

Generally a trend has emerged from the research, outlined in Figure 1, indicating a positive effect of participation in inquiry-oriented science activities on the broad skill area of language arts and reading. Guszak (1970) indicates that investigations of listening in educational research have been minimal compared to such research in language arts and reading.

It is the contention of the developers of the BSCS-ESSP program that children's experiences in learning are not naturally compartmentalized. Consequently, the developers have advocated the teaching of science, not only for scientific content and process skills, but also as a vehicle for the training of other fundamental areas of skill, notably those in the language arts. Specifically identified as a skill in language arts that is enhanced by participation in certain activities in BSCS-ESSP is skill in listening. Listening, as used in the BSCS-ESSP program,

Figure 1

HIGHLIGHTS OF RESEARCH IN SCIENCE EDUCATION ON READING AND LANGUAGE ARTS SKILLS

SCIENCE EDUCATION RESEARCH STUDIES: READING READINESS AND ORAL LANGUAGE

Author/Date	Variables investigated	Program used, sample used	Evaluation instrument	Conclusions
Ayers, J. B., Mason, G. F. 1966	The effect of SAPA science experiences on changes in reading readiness	SAPA—Level A. Kindergarten in Atlanta, Georgia	Metropolitan Readiness Test	Total reading readiness scores significantly higher in science groups. Listening, Numbers, and Copying scores higher in science group. Control group better on Word Meaning.
Rowe, M.B. 1968	The effect of SCIS on the verbalizing of innercity children	SCIS—Material Objects. First graders in New York City	Evaluation by trained observers	Observers not able to identify low verbal children in SCIS activities. The most spontaneous language (complete sentences) occurred during the SCIS activities, 200 percent more than language arts.
Ritz, W. C. 1969	The effect of SAPA and/or the Frostig Program of Visual Perception programs on reading readiness perceptual skills, and science process skills	SAPA—Level A; Frostig Program of Visual Perception. Kindergarten in New York and Pennsylvania	Metropolitan Readiness Test; Frostig Developmental Test of Visual Perception; Science Competency	No significant differences between groups on reading readiness. Effects from the Frostig program; only short term science competency changes.
Neuman, D. B. 1970	The effect of science experiences, written by the author, on reading readiness	The author's activities were based on SCIS and SAPA	Metropolitan Reading Readiness Test	Children having science activities scored significantly better on reading readiness. They did better on Listening, Numbers, Vocabulary, and Alphabet; control group did better on Matching.
Kellogg, D. H. 1971	The effect of SCIS—Material Objects unit on reading readiness was compared with Harper & Row Reading Readiness Program	SCIS—Material Objects. First graders in Ada, Oklahoma	Metropolitan Reading Readiness Test	SCIS group outscored control group on total reading readiness score; particularly better on Word Meaning, Listening, Matching, Alphabet, and Numbers. Control did better on Copying.
Huff, P., Languis, M. 1973	The effect of SAPA on oral communication skills of disadvantaged children, including speaking and listening	SAPA—Level A. Kindergarten in Columbus, Ohio	Test of Oral Communication Skills (TOCS)	Experimental group significantly better on oral transmission skills, but not listening comprehension receiving subskill.
Bethel, L. J. 1973	The effect of science experiences, written by the author, on oral communication and classification skills	Author produced. Third graders in Philadelphia innercity	Test of Oral Communication Skills; Goldstein-Scheerer Object Sorting Test	Experimental science group showed significantly improved classification and transmitting and receiving oral communication skills.

---

is defined as the total process of attending to, recognizing, and interpreting spoken symbols. Little empirical data is available however, to support this contention. Therefore, the problem to be investigated was whether participation in an inquiry program does enhance listening skills.

PURPOSE

The purpose of this study was to investigate the effects of participation in selected activities from experimental materials of the BSCS-ESSP on the enhancement of skills in listening of first-grade children. Data from four independent variables — gender, age, previous experience in school and socio-economic level — were included in this study.

The dependent variable was the total performance score on the Cooperative Primary Series Listening Test (CPSLT). Performance on the instrument, *Science* was also examined as a measure of science competency.

RESEARCH QUESTIONS

The study was designed to investigate the following research questions:

1. Does participation in selected experimental activities from the BSCS-ESSP enhance skills in listening as measured by total performance scores on the CPSLT? (Educational Testing Service, 1967.)
2. Does participation in selected experimental activities from the BSCS-ESSP enhance the following specific listing skills:
    a. Identification
    b. Recall-Remember
    c. Interpretation-Evaluation-Inference as measured by CPSLT?
3. Do the following student-related factors affect achievement in listening skills as measured by total performance scores on the CPSLT?
    a. Age b. Gender c. Previous experience in school
4. Does participation in selected experimental activities from the BSCS-ESSP enhance cognitive skills in science as measurd by performance scores on the test, *Science*?
5. Do the following student related factors affect cognitive skills in science as measured by performance scores on the test, *Science*?
    a. Age b. Gender c. Previous experience in school

RESEARCH DESIGN

SAMPLE

The study was conducted in a central Texas medium-sized school district. Random selection was used to identifiy six participating schools and nineteen teachers. Four hundred and eighty-one children in the first grade participated in the study.

TREATMENT

The treatment consisted of participation in twelve activities selected from the experimental version of BSCS-ESSP. The BSCS-ESSP is a multidisciplinary, hands-on, inquiry-oriented program that recognizes the importance of listening skills in the elementary school classroom. The selected activities were screened for instructional components related to developing listening skills and included group discussions involving questioning strategies, audiotape and other media-related activities and concepts related to sound. Teachers were requested to follow closely the outlined dialogue in the discussion format provided with each lesson. A schedule of instruction, including suggested time limits for each lesson, was provided to standardize procedures.

EXPERIMENTAL DESIGN

The Solomon Four-Group design was employed. Pre- and posttests were performed using CPSLT and *Science*. The unit of analysis was individual subjects.

Analysis of covariance (Veldman, 1974) was performed on those sets of data that included both pre- and posttest information; the pretest was treated as the covariate and the posttest as the criterion variable. For those subjects without pretest information, analysis of variance comparisons between groups was made. Analysis of variance (Veldman, 1974) was also used to compare subjects according to the independent variables.

INSTRUMENTATION

Alpha reliability for *Listening* Form 12B was calculated at r=.81 (ETS, 1976). Reviewers in Buros (1972; 1978) stated that the CPSLT does measure basic verbal skills at an application level. In addition, Airasian (1978) stated that the comparison of test specifications and the test item on the CPSLT indicates that the content validity of all tests is excellent.

Science, a competency instrument, was excerpted from the BSCS-ESSP trial version. This modified version was piloted and an alpha reliability of r=.84 was computed for participants in the experimental group.

RESULTS

For the first question investigated, concerning the enhancement of listening skills among subjects in the experimental group, it was found that participation in selected activities from the BSCS-ESSP does enhance total skill in listening for those subjects who achieved pretest scores at or above the group mean. These subjects displayed the greatest total performance scores on *Listening*. However, there was no significant difference between the experimental and control groups (see Table 1).

The second question investigated was the enhancement of component skills in listening-identification, recall-remember and interpretation-evaluation-inference. Results revealed that participation in BSCS-ESSP does not enhance

Table 1

ANALYSIS OF VARIANCE FOR BETWEEN GROUPS COMPARISON ON TOTAL POSTTEST PERFORMANCE SCORES ON *LISTENING*

S. S.	df	M. S.	F	P
36.2744	1	36.2744	.6724	.4180

	N	Posttest Mean	S. D.
Experimental	242	34.9752	7.9885
Control	192	35.5573	6.4416

the subskill of identification. But it does enhance the subskills recall-remember and interpretation-evaluation-inference.

The third question investigated student-related factors and their effect on achievement in listening skill as measured by total performance scores on CPSLT. Analysis of variance was employed to compare demographically sorted groups according to total performance scores on *Listening*. Table 2 contains a summary of the findings. The results reveal that age and previous school experience do have a significant effect of listening scores in the experimental group when compared to the control group.

Table 2

SUMMARY OF COMPARISON OF LISTENING WITH DEMOGRAPHIC VARIABLES: TOTAL LISTENING.

Demographic Variables	s.s.	df	M.S.	F	P
Age	1045.58	5	209.12	4.18	.00*
Sex	135.18	1	135.18	2.52	.11
Previous experience	1633.73	4	408.43	8.34	.00*

*Significant at the 0.01 level

The fourth question dealt with whether participation in the BSCS-ESSP enhanced cognitive skills in science. Analysis of the data revealed that the science activities did not have a significant effect of the cognitive skills of the experimental group when compared with the control group.

For question five, differential effects were seen with regard to some demographic variables (age, sex, previous experience) investigated on comparisons of performance scores on *Science*. Age and previous experience in school did have an effect in the experimental group when compared to the control group. Gender of the subjects was not seen to be a factor in any comparisons made (see Table 3).

Table 3

SUMMARY OF COMPARISON OF COGNITIVE SKILLS IN SCIENCE WITH DEMOGRAPHIC VARIABLES: SCIENCE.

Demographic Variables	s.s.	df	M.S.	F	P
Age	237.89	5	47.57	3.38	.01*
Sex	5.81	1	5.81	.385	.54
Previous experience	265.25	4	66.31	4.74	.00*

*Significant at the 0.01 level

SUMMARY

Participation in selected activities from the BSCS-ESSP has differential effects with respect to the skills in listening and skills in science. Members of the experimental population show gains in areas of skill in listening and cognitive skills in science.

Participation in the program does affect enhancement of the subskill recall-remember for those experimental subjects who entered the program with performance scores on the pretest that were above the mean. The listening subskill, interpretation-evaluation-inference, also is enhanced by participation in BSCS-ESSP for those subjects who also scored at or above the mean on the pretest.

The student-related factors of age and previous school experience affect total performance scores on *Listening* and *Science*. Gender was not shown to be a factor (see Tables 2 and 3). Performance scores of those subjects with more experience in school were greater on *Listening*. The oldest and youngest children consistently had lower performance scores on *Listening* and *Science;* some variation among the subjects was found in the middle age range. Those subjects for whom first grade was the first experience in school attained the lowest performance scores.

The student-related factors of age and previous experience in school were all shown to have an effect on the three subskills of *Listening* — identification, recall-remember and interpretation-evaluation-inference. Gender was not shown to be a factor in these skills. Age was a significant factor in all listening performance scores on the subskills of *Listening* for comparisons between the experimental and control subjects. Previous school experience was shown to be a significant factor in performance scores. Those subjects with more experience in school were more successful than those with less or no previous experience in school.

DISCUSSION

The overall findings support the position that participation in selected activities from the BSCS-ESSP enhance skills in listening for at least some first-grade students. Similarly, the program has differential effects with respect to specific subskills in listening. The segment of population that benefitted from participation includes those students who had an entering skill level at or above the mean. A differential effect was also seen in the enhancement of cognitive skills in science. Cognitive skills in science showed enhancement only for a segment of students from lower socioeconomic levels.

The overall finding that a hands-on, multidisciplinary science program such as BSCS-ESSP can enhance skills associated with language arts, and specifically listening, are compatible with the research of Ayers & Mason (1966), Rowe (1968), and Neuman (1970). This finding further supports the work of Pronovost (1959) and Czekanski (1974), advocates of an instructional approach that integrates instruction in listening with instruction in science.

The overall finding that lower socioeconomic level students' listening comprehension was not enhanced by participation in an inquiry-oriented and activity-oriented science program is consistent with the findings of Huff and Languis (1973). Their study, performed with kindergarten students from the lower socioeconomic level who participated in Science-A Process Approach activities, did not indicate an enhancement of listening ability, but did indicate an enhancement of speaking skills. Certainly speaking skills are considered a precursor to other language abilities. Other studies (Ayers & Mason, 1966; Neuman, 1970) did find that the listening skills of young children were enhanced by participation in science programs, but listening in these studies was evaluated as a component of reading readiness.

While gains were found for concepts in science for a segment of the experimental population, overall findings indicate an analysis of the level of difficulty of BSCS-ESSP materials for use in a particular classroom or grade level would be recommended. Neuman (1970), in previous research (See Table 1) only found short term gains in science competency. That children from lower socioeconomic levels demonstrated benefits from participation in a hands-on, inquiry-oriented program is consistent with other research outlined in Table 1. Implications of such instructional benefits provide support for early intervention and remedial programs such as Head Start and Title 1.

The focus of this study has been the enhancement of listening skills. The study indicates that those students who possess the requisite ability are able to benefit from a science program that is hands-on, activity-centered and inquiry-oriented with respect to gains in listening skills. Enhancement of basic concepts in science for students from lower socioeconomic backgrounds is also indicated. The coupling of these two findings have practical implications for elementary school curricula, namely, that science should be an integral part of all elementary education programs.

REFERENCES

Airasian, P. W. Review of cooperative primary tests. *In the eight mental measurements yearbook.* Highland Park, N. J.: Gryphon Press, 1978.

Ayers, J. B. and Mason, G. E. Differential effects of science — a process approach upon change in metropolitan readiness test scores among kindergarten children. *Reading Teacher, 22,* 435-58, April, 1966.

Barufaldi, J. P. and Swift, J. R. W. Children learning to read should experience science. *Reading Teacher,* 1977, 30, 338-393.

Bethel, L. J. Science inquiry and the development of classification and oral communication skills in innercity children. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, Pa., 1974.

Biological Sciences Curriculum Study — Elementary School Sciences Program (BSCS-ESSP). The BSCS elementary school science program — A status report. *BSCS Newsletter,* No. 54: 6-18, February, 1974.

Butts, D. *Teaching science in the elementary school.* New York: The Free Press, 1973.

Buros, O. K. *The seventh mental measurements yearbook.* Highland Park, N. J.: Gryphon Press, 1972.

Czekanski, E. The neglected scientific skill: Listening. *Science and Children, 12,* 1974.

Duker, S. Listening. *Listening: Readings.* Duker (Ed.), New York: Scarecrow Press, Inc., 1966, 165.

Educational Testing Service. *Cooperative primary tests handbook.* Princeton, New Jersey, 1967.

Guszak, F. J. Diagnostic reading instruction in the elementary school. New York: Harper and Row, 1972.

Huff, P., & Languis, M. The effects of the use of activities of SAPA on the oral communication skills of disadvantaged kindergarten children. *Journal of Research in Science Teaching, 10,* (2), 1973, 165.

Kellogg, D. B. *An investigation of the effect of SCIS first year unit, material objects, on gains in reading readiness.* Unpublished doctoral dissertation, University of Oklahoma, 1971.

Neuman, D. B. *Effect of kindergarten science experiences of reading readiness.* Final report. Presented: National Association for Research in Science Teaching, Pasadena, California, 1969. ERIC EDC 044 264, 1970.

Neuman, D. B. Teachers don't have time NOT to teach science. *Science and Children, 12,* (3), Nov.-Dec., 1974.

Phillips, G. M., Dunham, R. E., Brubaker, R. and Butt, D. *The development of oral communication in the classroom.* New York: Bobbs-Merrill Publisher, 1970.

Pronovost, W. *The teaching of speaking and listening in the elementary school.* New York: Longmans, Green and Co., 1959.

Ritz, W. C. *The effect of two instructional programs (Science — A Process Approach and the Frostig Program for the Development of Visual Perception) on the attainment of reading readiness, visual perception, and science process skills in kindergarten children.* Unpublished doctoral dissertation, State University of New York, Buffalo. University Microfilms, Inc., Ann Arbor, Michigan, 1969.

Rowe, M. B. SCIS in the inner city school. *Science Curriculum Improvement Study Newsletter, 11,* 6-7, 1968.

Rowe, M. B. Help is denied to those in need. *Science and Children. 12,* (6), 23-25, March, 1975.

Stauffer, R. G. Reading as cognitive functioning. *Theoretical models and processes of reading.* H. Singer and R. B. Ruddell (Eds.), International Reading Association, Newark, Delaware, 1970.

Veldman, D. J. *V-Stat user manual.* Austin, Texas: The University of Texas, 1974.

ON THE EFFECTIVENESS OF TEACHER-PREPARED HANDOUTS IN LEARNING SCIENCE CONCEPTS

JAMES A PETRICH
DEPARTMENT OF CHEMISTRY
SAN ANTONIO COLLEGE
SAN ANTONIO, TEXAS

INTRODUCTION

A greater percentage of class time in college science courses is devoted to lectures than to any other form of instruction (Gagne, 1970, 370). Although it has its limitations, the lecture is generally regarded as the most efficient method of instruction. It is usually effective and is always readily available. Nevertheless, the thought that the lecture method could be improved upon lingers. A lecture does convey the desired information and, perhaps, even helps to mold a student's attitudes, but without student interaction it does not facilitate retention (Schumacher, Liebert & Fass, 1975; Howe, 1976). Therefore, rather than replacing the lecture method with any other method, some teachers have sought to augment the lecture by supplying the student with lecture notes (Klemm, 1976).

PROBLEM

This investigator studies the effectiveness of teacher-prepared handouts by comparing the achievement of students who were not supplied them with that of students who were supplied with two different kinds: transcriptions and outlines. It was believed that these two types of notes represent extremes of organization, accuracy of detail and, particularly, cognitive involvement of the student. This involvement was seen as a necessary aspect of learning that is largely missing when the lecture method is used. The organization of the scripts was no more apparent than that of the lecture itself, but the outlines presented an immediately apparent organization of the lecture contents so that the student did not have to generate his own. These lecture aids also differed from note-taking and from each other in the amount of detail provided and, therefore, the accuracy of the details as received by the student. The scripts suggested to the student that everything being said was already recorded, so no action by him was necessary. However, the outlines, which had plenty of space for student notes, encouraged the student to pay attention and fill in the details while still giving him enough time to interpret and assimilate the oral material.

PURPOSE

The purpose of this study was to determine the influence of organization and accuracy of detail on cognitive involvement as measured by achievement. Three primary research questions were examined in this investigation:

1. Is learning directly proportional to the extent with which a student is cognitively involved in a learning situation?

2. Is learning directly proportional to the level of organization of teaching material?
3. Is learning directly proportional to the accuracy of the details in teaching materials, as received by the student?

RESEARCH DESIGN

PROCEDURE

Three classes of chemistry students at a central Texas college were selected as subjects. The course was the first semester of freshman chemistry for science majors. The classes consisted of 18, 19 and 17 students, respectively, although 1 member of the third class did not complete the first topic and one member of each class did not complete the third topic. The subjects were 70% male and had an average age of 20.7, ranging from 17 to 31 years old. The primary language was English for 87% of the subjects. Their ethnic background was mostly European, non-Spanish (51%); Hispanic (37%) or Iranian (8%).

Three topics, which differ in the nature of their content, were selected for three distinct investigations:

Topic 1 — ACIDS, BASES AND SALTS, an introduction to Arrhenius-type acids and bases, a topic which was largely descriptive but also somewhat theoretical, mathematical and experimental, because a large number of demonstrations were presented;

Topic 2 — OTHER TYPES OF ACIDS AND BASES, a description of the Bronsted-Lowry and Lewis theories of acid-base behavior, a topic which was highly theoretical; and

Topic 3 — THE EFFECT OF SOLUTES ON SOLVENTS, a discussion of colligative properties (vapor pressure depression, boiling-point elevation and freezing-point depression), a topic which was theoretical and highly mathematical.

In order to equalize the instruction among the three classes each lecture was videotaped. The lectures were 35, 20, and 25 minutes long, respectively.

The classes were randomly designated as NO AIDS, SCRIPT and OUTLINE, receiving no handouts, a script of each lecture or an outline of each lecture respectively. Students in all three classes were first administered a pretest on prerequisite knowledge and skills required to successfully learn the three topics. Then they were instructed to read the relevant pages from the textbook for the first topic and the students in the SCRIPT and OUTLINE classes were given their notes. The next day the first videotaped lecture was presented. Finally, the students were tested on the lecture on the third day. Except for the pretest, this procedure was then repeated for the second and third topics.

INSTRUMENTATION

The pretest and the achievement tests on the three topics consisted of five-alternative multiple-choice questions. The pretest had 44 questions, and each of the achievement tests had 20 questions. However, some of these questions were omitted from the analysis in order to improve the reliability of the tests, leaving 37 questions on the pretest and 18 questions on each of the achievement tests. The reliabilities (Kudor-Richardson procedure) on the pretest and the three achievement tests were 0.84, 0.73, 0.78 and 0.89 respectively. Further details on the composition of the classes, the content of the handouts, the preparation of the tests, the content of the tests, and the test results are described elsewhere (Petrich, 1979).

RESULTS

Neither the students' college entrance examination scores (ACT and SAT) nor their grades on the pretest indicated any statistically significant differences among the classes in ability or prerequisite knowledge and skills. The average scores on the pretest were 61, 58, and 56 for the NO AIDS, SCRIPT and OUTLINE classes, respectively. The average scores on the achievement tests for each class on each topic are given in Table 1.

Table 1

MEANS OF TREATMENT GROUPS ON ACHIEVEMENT TESTS: PERCENT CORRECT

	Topic 1	Topic 2	Topic 3
NO AIDS	78	61	77
SCRIPT	70	45	61
OUTLINE	62	32	51

These results were examined by analysis of variance and by analysis of covariance, using either ACT score or pretest score as the covariate. These analyses indicated a statistically significant difference between the NO AIDS and OUTLINE classes on the first topic, among all three groups on the second topic, and between the NO AIDS class and either the SCRIPT class or the OUTLINE class, depending upon whether or not the ACT score was used as a covariate, on the third topic. All of these differences had a probability of 0.05 or less of happening by chance (Table 2). For more details, see Petrich and Montague (1981).

DISCUSSION

Subjects who took notes without lecture aids scored significantly higher than those who received an outline of the lecture on both the descriptive and the

Table 2

STATISTICAL SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS

	Topic 1		Topic 2		Topic 3	
Covariate	None	Relevant Pretest Questions	None	ACT	None	ACT
NO AIDS vs. SCRIPT	.05	.05	.01	.05	.05	.05
NO AIDS vs. OUTLINE	.01	.01	.01	.01	.01	.05
SCRIPT vs. OUTLINE	.05	.05	.01	.05	.05	.05

theoretical topics. However, the difference between the groups on the third topic, which was more mathematical, was not statistically significant when the initial difference between the groups, as indicated by their ACT scores, was considered. Furthermore, on the second topic, which was highly theoretical, subjects who were not given lecture aids scored significantly higher than those who received a script of the lecture, and those who were given a script scored significantly higher than those given an outline.

The superior performance of the subjects who received no lecture aids indicated that a high level of cognitive involvement was produced by note-taking. These subjects were given the same amount of organization as the subjects who received the scripts of the lectures and less organization than the subjects who were given outlines. Furthermore, their recall of the lectures was subject to more inaccuracies from writing down the details than the recall of the subjects who were given a script or an outline, although the effect of this influence was, no doubt, related to the cognitive level of the test questions.

The intermediate position of the script in facilitating learning was surprising, considering the presumed importance of organization (Bigge, 1976, 261). However, Schumacher et al (1975) did find that a text with little or no apparent organization facilitated learning because it required the student to organize the information himself and, thus, become involved with it. Furthermore, his organization may be a better organization for him than an organization given to him by the teacher. Apparently, this is how the scripts were used by the students in this study. Also, the script did provide complete accuracy of detail on the information provided in the lecture.

The outlines apparently did not provide enough cognitive involvement. Perhaps they were so detailed that the student assumed that everything he needed was provided. Furthermore, the organization of the outlines was not necessarily the organization he would have used if he had prepared his own outlines.

CONCLUSIONS

The findings of this study suggest that the distribution of either scripts or outlines of lectures should not be considered as a general cure for the deficiencies of the lecture method of instruction. In fact, this study did not determine any type of material for which such aids would be beneficial. However, this study did suggest that they are particularly undesirable when the student is learning highly theoretical concepts, such as in Topic 2 (OTHER TYPES OF ACIDS AND BASES). Even when the material is primarily mathematical or descriptive, such as in Topic 1 (ACIDS, BASES and SALTS) and Topic 3 (THE EFFECT OF SOLUTES AND SOLVENTS), taking notes without any handout appears to induce more learning than using instructor-prepared lecture notes. Apparently a student provided with instructor aids for recording and organizing lectures is not as cognitively involved in interpreting and assimiliating the material as a student left to record and organize it on his own.

REFERENCES

Bigge, M. L. *Learning theories for teachers.* (3rd ed.) New York: Harper & Row, 1976.

Gagne, R. M. *The conditions of learning.* (2nd ed.) New York: Holt, Rinehart & Winston, 1970.

Howe, M. J. A. What is the value of taking notes? *Improving College and University Teaching*, 1976, *24*, 22-24.

Klemm, W. R. Efficiency of handout "skeleton" notes on student learning, *Improving College and University Teaching*, 1976, *24*, 10-12.

Petrich, J. A. *The effect of instructor-prepared handout materials on learning from lecture instruction.* Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas. Ann Arbor, Michigan: University Microfilms, 1979, No. 7920189.

Petrich, J. A. & Montague, E. J. The effect of instructor-prepared handout materials on learning from lecture instruction. *Journal of Research in Science Teaching*, 1981, *18, 177-187.*

Schumacher, G. M., Liebert, D. & Fass, W. Textual organization, advance organizers and the retention of prose material. *Journal of Reading Behavior*, 1975, *7*, 173-180.